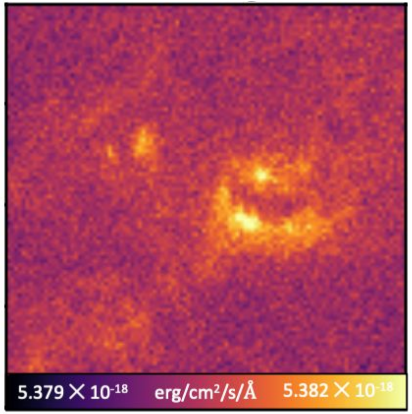
Merging galaxies in *HST* and *JWST*:

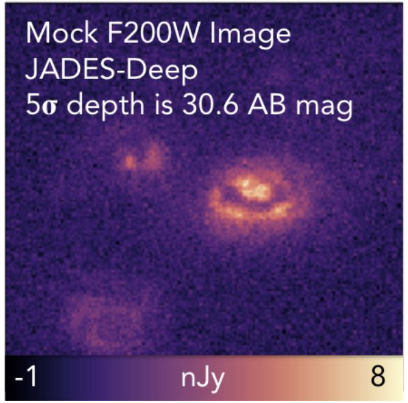
An interpretable suite of CNNs for identifying and understanding merger features from

cosmic coffee hour ……. to ……. cosmic brunch

HST F814W

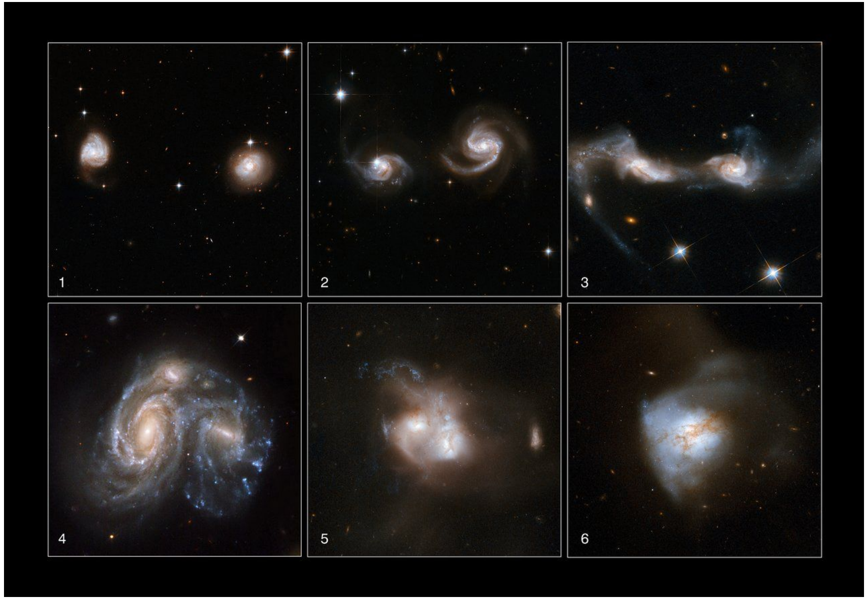


JWST F200W



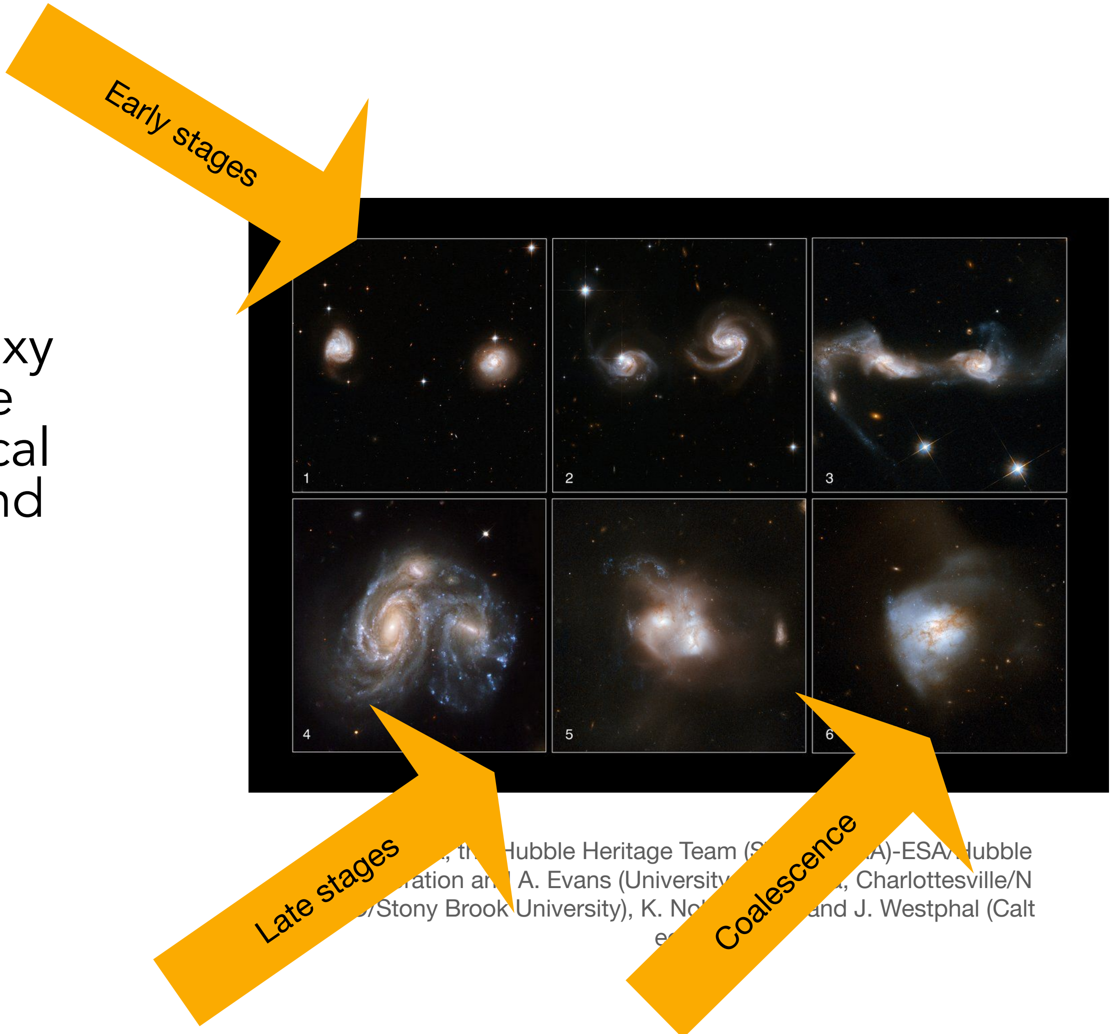
Aimee Schechter and Becky Nevin

Mergers can alter galaxy morphologies, provide evidence for hierarchical structure formation, and turn on AGN and star formation



NASA, ESA, the Hubble Heritage Team (STScI/AURA)-ESA/Hubble Collaboration and A. Evans (University of Virginia, Charlottesville/NRAO/Stony Brook University), K. Noll (STScI), and J. Westphal (Caltech)

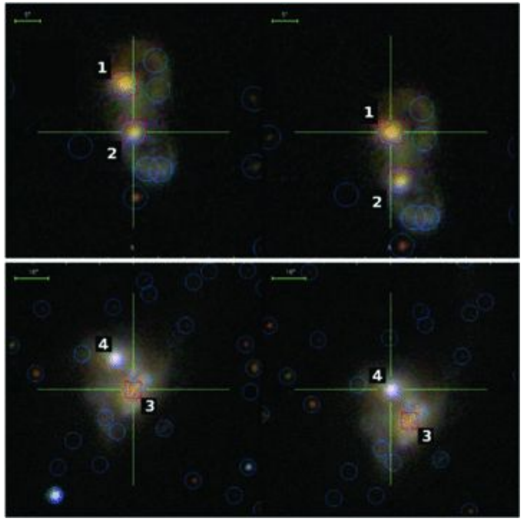
Mergers can alter galaxy morphologies, provide evidence for hierarchical structure formation, and turn on AGN and star formation



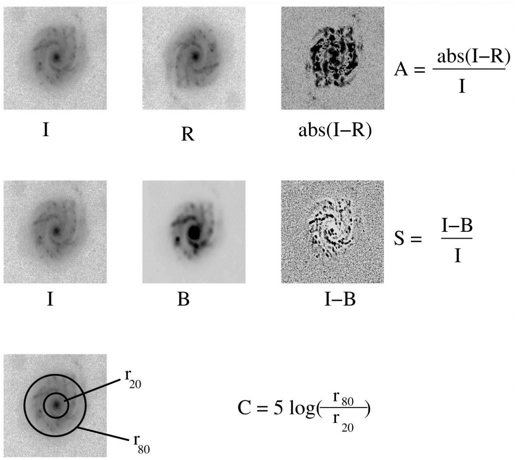
, th Hubble Heritage Team (S A)-ESA/Hubble
ration an A. Evans (University , Charlottesville/N
/Stony Brook University), K. No and J. Westphal (Calt
e

Mergers have been identified visually and quantitatively in the past

Citizen Scientists identify mergers visually through the Galaxy Zoo projects (e.g., Darg et al. 2010)



Quantitative measurements such as Concentration, Asymmetry, Clumpiness, and measures of light distribution (e.g., Concelise 2003, Lotz et al. 2004)

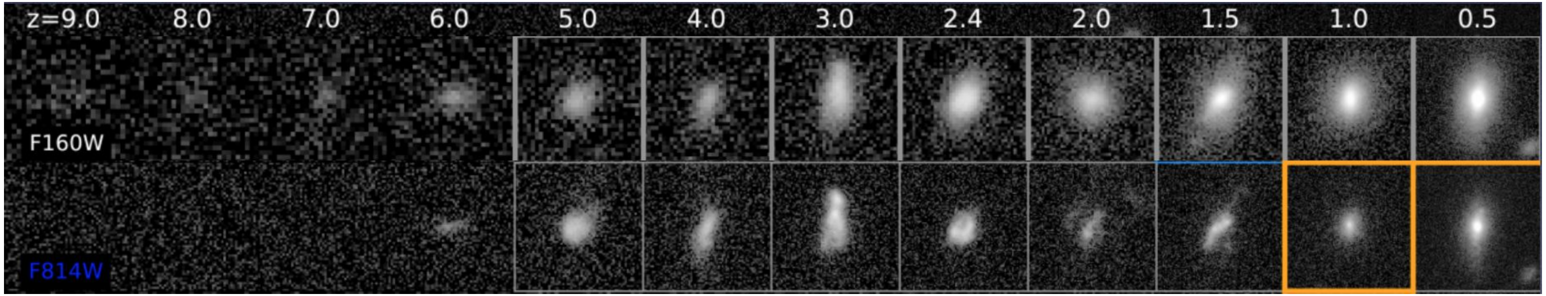


Machine Learning can recognize more merger stages, and handle large data sets

- Snyder et al. 2019 used a random forest classifier on Illustris *HST* mock images

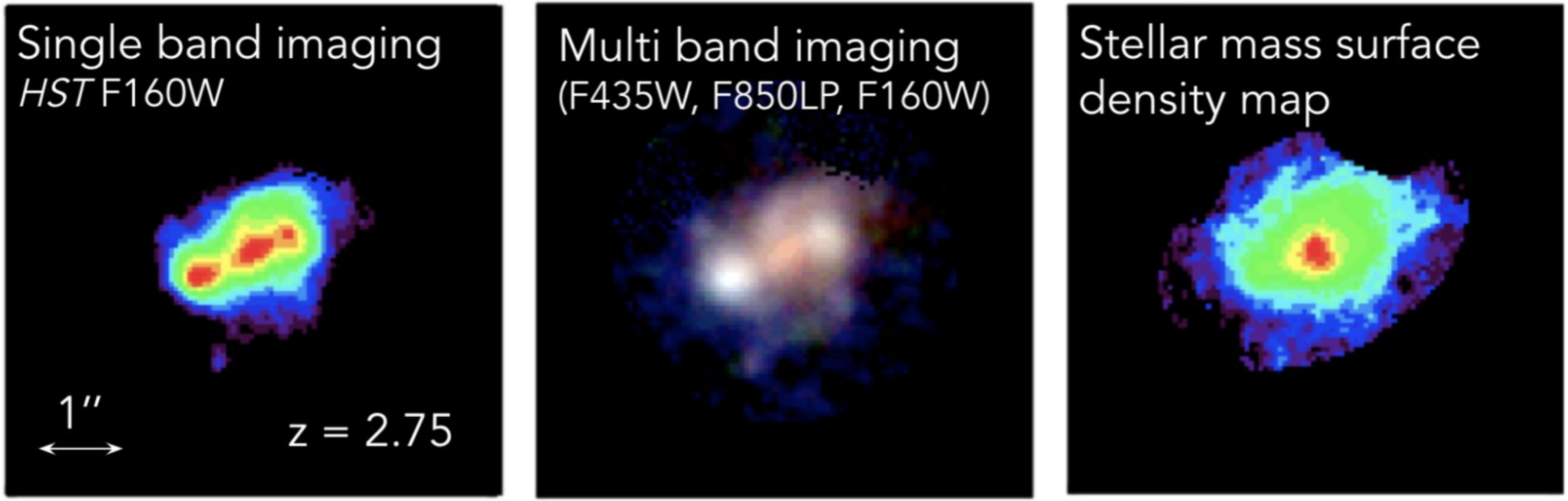
- Bottrell et al. 2019 used convolutional neural networks for merger classification and discusses important aspects of mock images

- Ferreira et al. 2020 identified mergers and calculated a merger rate with mock CANDELS images from IllustrisTNG300



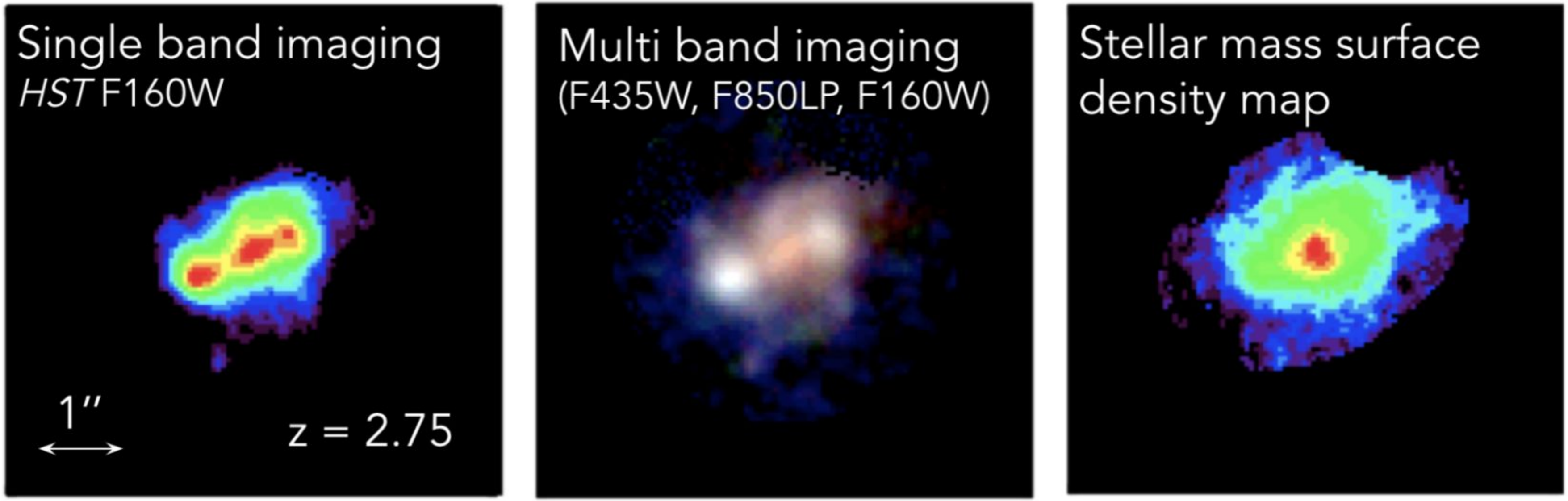
Snyder et al. 2019

High redshift galaxies are inherently clumpy and mergers are harder to identify



Cibinel+2015

Tools derived from multiple filters can enable more accurate merger identification



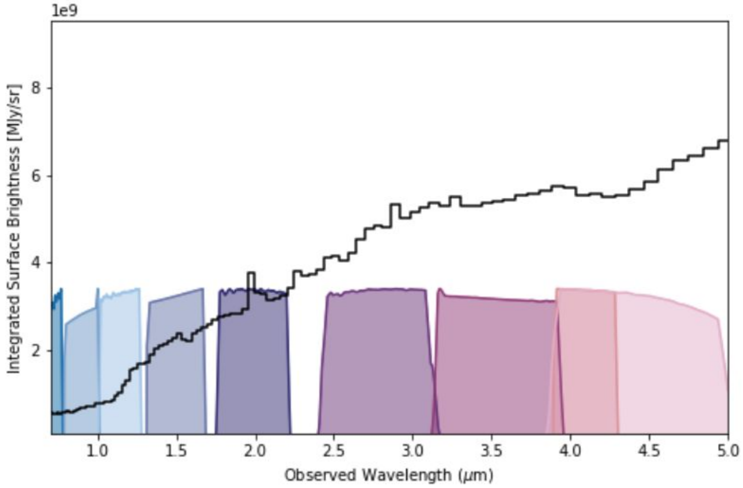
Cibinel+2015

CANDELS is great for studying mergers

- *HST* CANDELS has high spatial resolution images in optical and infrared filters
- Redshift range covers the peak of galaxy assembly (we use $0.2 < z < 3$)
- Investigate connection between merger classification/stage, AGN activity, and star formation

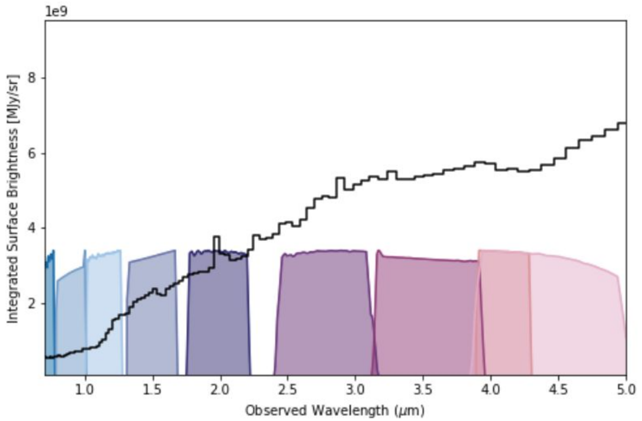
JWST is great for studying high redshift mergers

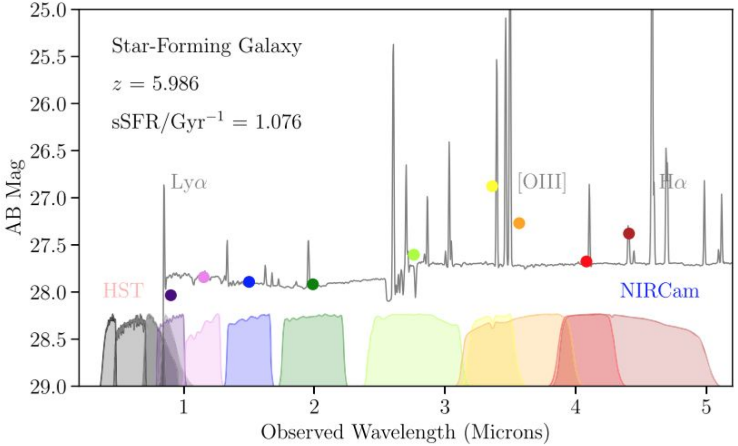
- Deep surveys such as JADES will give us a window into high-z galaxy morphologies currently inaccessible to HST (0.3kpc at z = 3)
- Role of minor/major mergers in driving mass growth in the early universe, specifically of massive compact ellipticals
- Role of mergers in disk instabilities



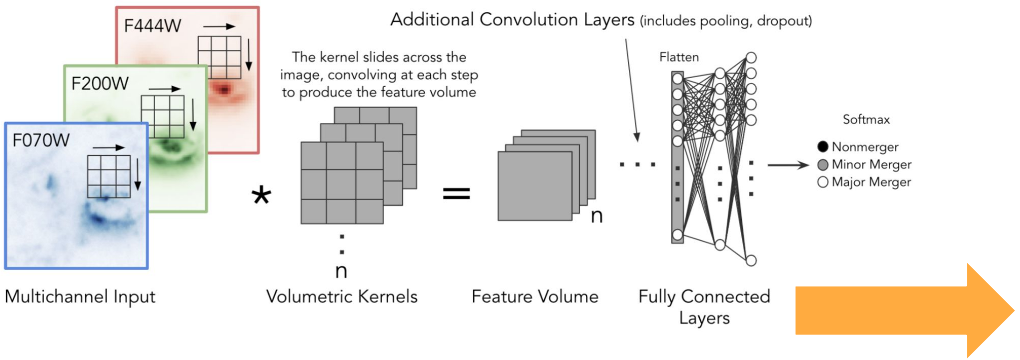
JWST is great for studying high redshift mergers

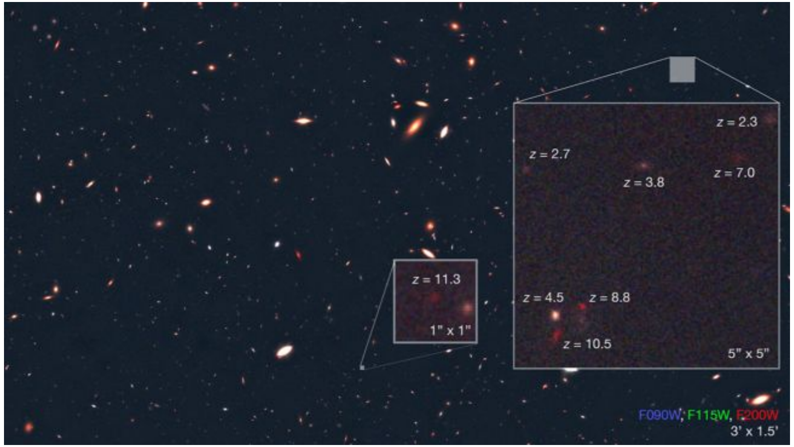
- Deep surveys such as JADES will give us a window into high-z galaxy morphologies currently inaccessible to HST (0.3kpc at z = 3)
- Role of minor/major mergers in driving mass growth in the early universe, specifically of massive compact ellipticals
- Role of mergers in disk instabilities
- Follow-up spectroscopic observations from GTO and ERS surveys





Joint talk journey

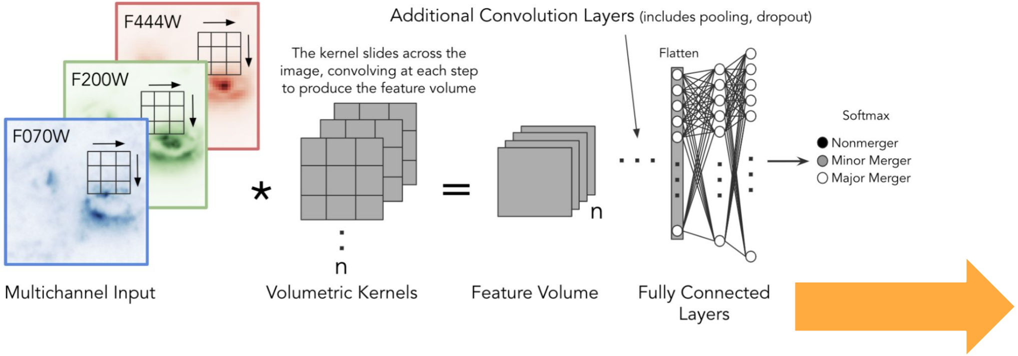


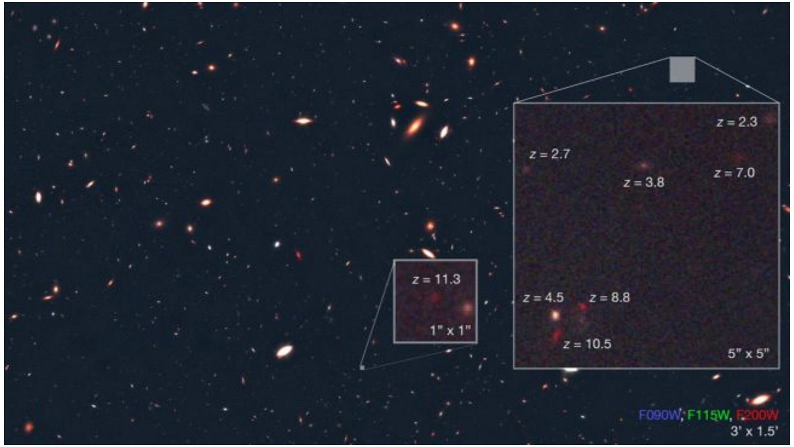


Guitarra image from Williams+2018

1) Build and train suites of CNNs
2) Interpret CNNs (identify merger features across cosmic time)
3) Use domain adaptation to classify *HST* and *JWST* fields

CiNNamonroll: A convolutional neural network framework to identify mergers in *JWST*

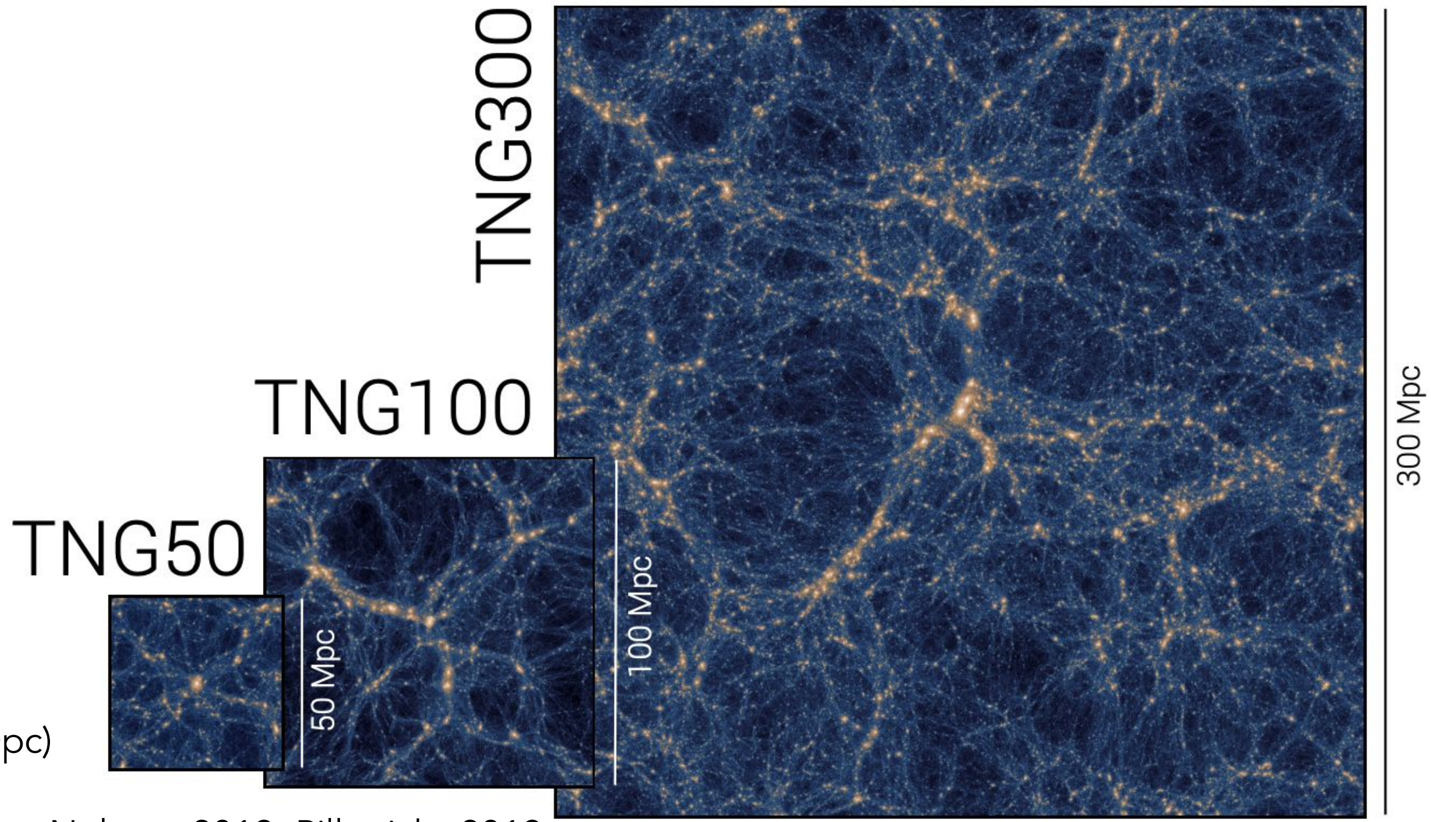




Guitarra image from Williams+2018

1) Build and train suites of CNNs
2) Interpret CNNs (identify merger features across cosmic time)
3) Use domain adaptation to classify *HST* and *JWST* fields

Training set! → Illustris TNG50

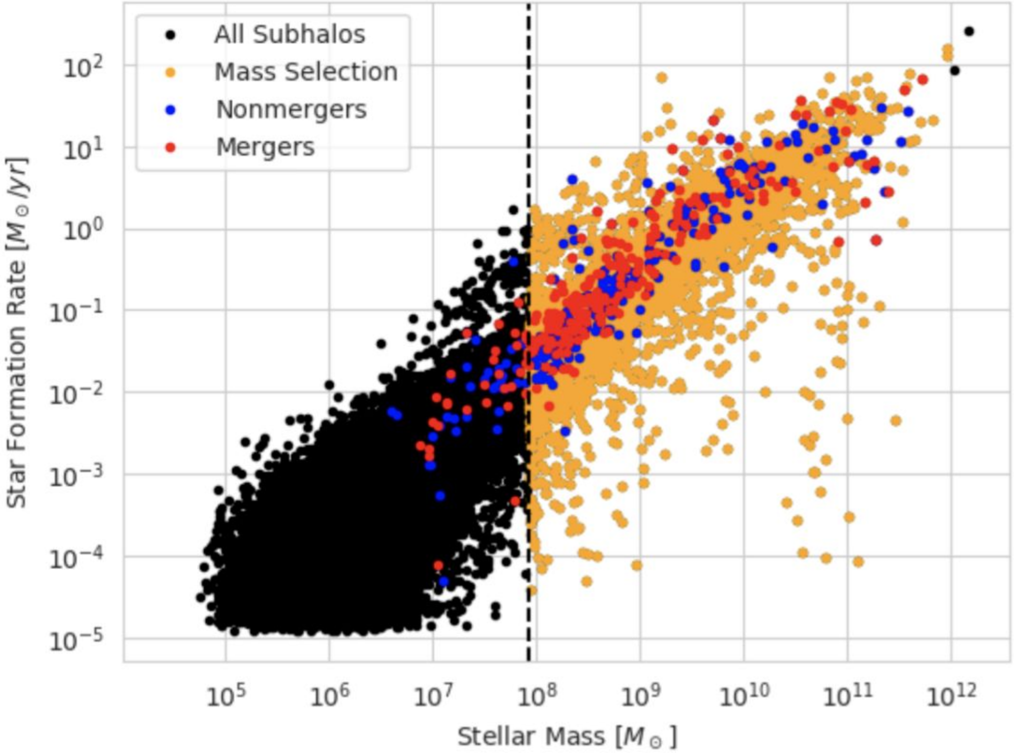


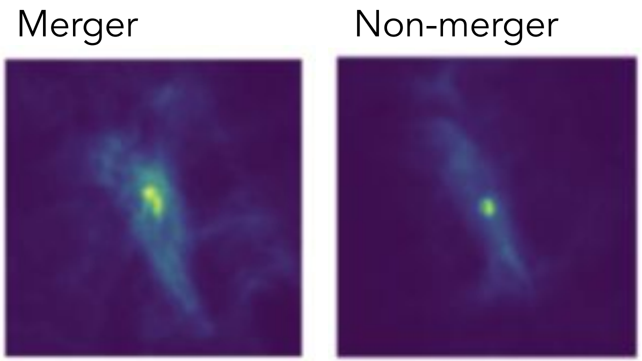
~72pc resolution
(TNG100 is about ~190pc)

TNG50 presentation papers: Nelson+2019, Pillepich+2019

Identify merging and nonmerging galaxies in TNG50

There are ~300 merging galaxies for z=1

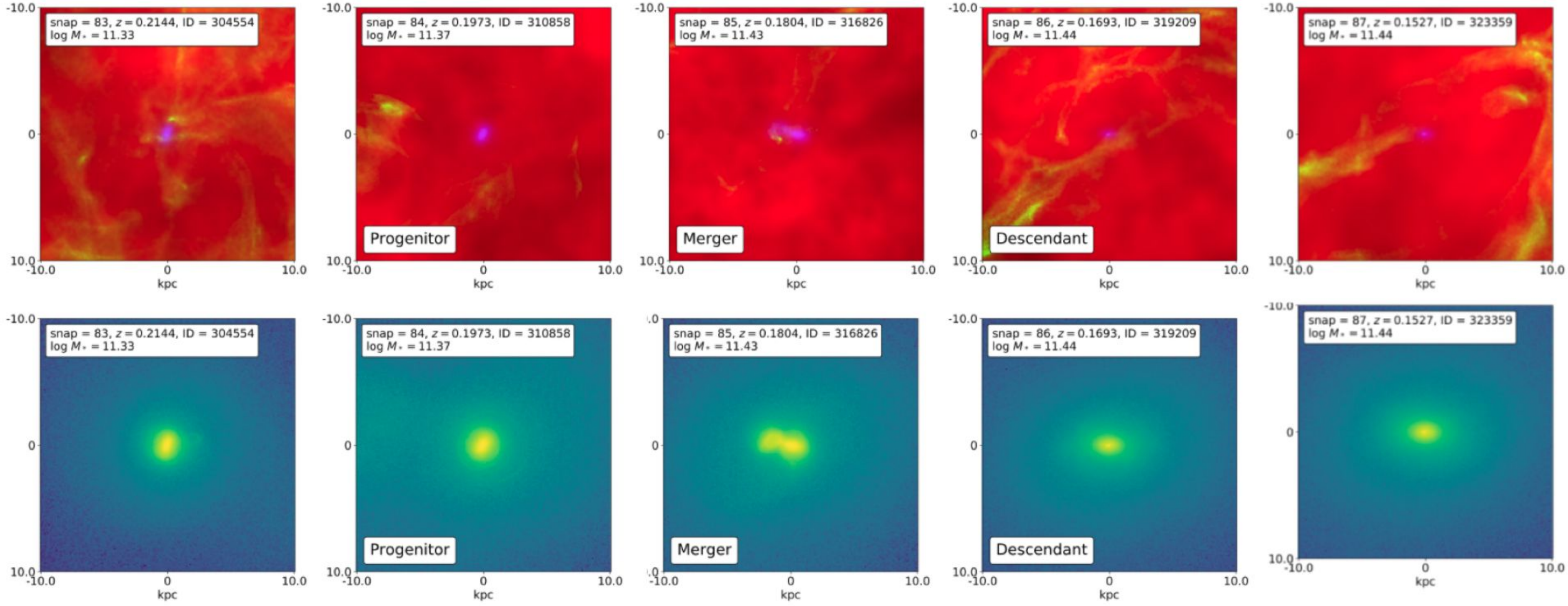




Gas density

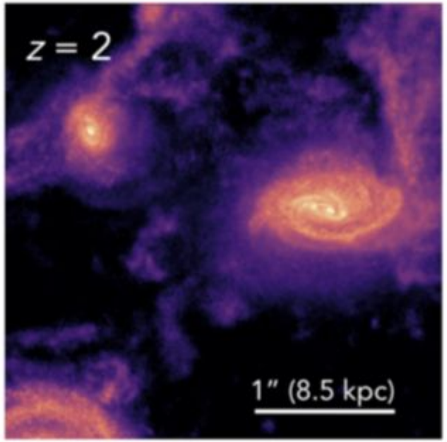
Particle maps are three color images (stars, gas, metals)

Major merger ($\mu_* = 0.41$) progenitor at z=0.2



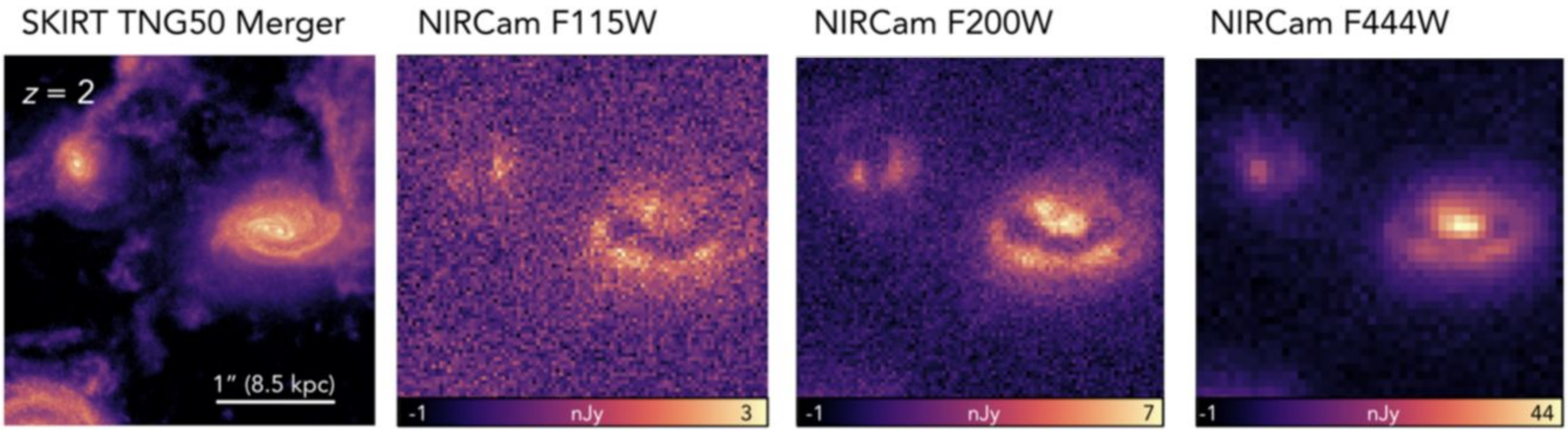
To create realistic mock images, we run SKIRT radiative transfer on the full sample of mergers and non-mergers

SKIRT TNG50 Merger



Jacob Shen

The final step is to create observationally realistic images by introducing noise, background sources, and instrument effects

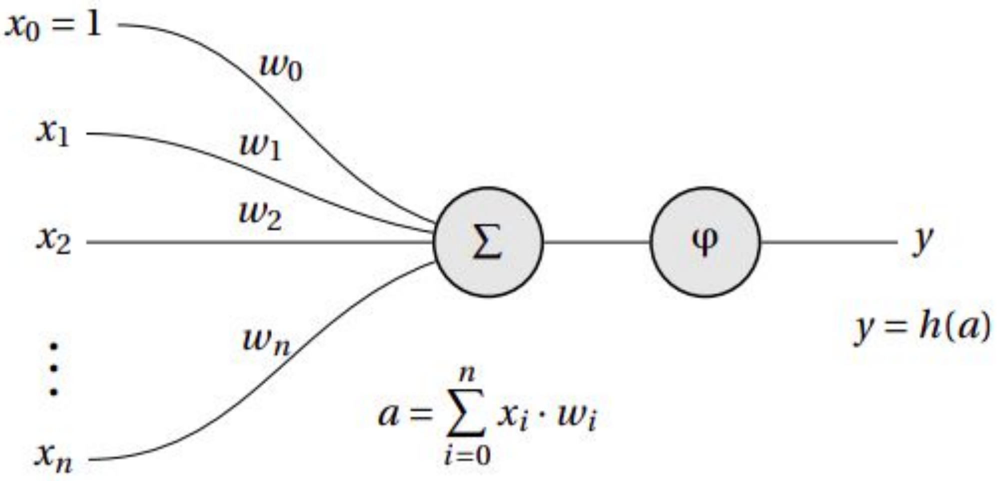


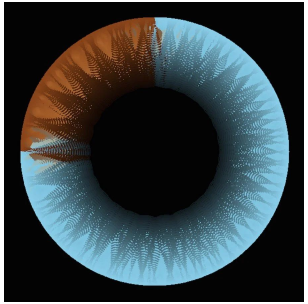
Discuss:

What is the best way to add realistic background galaxies to these images? Masking central galaxies or placing in a real field where there are no galaxies?

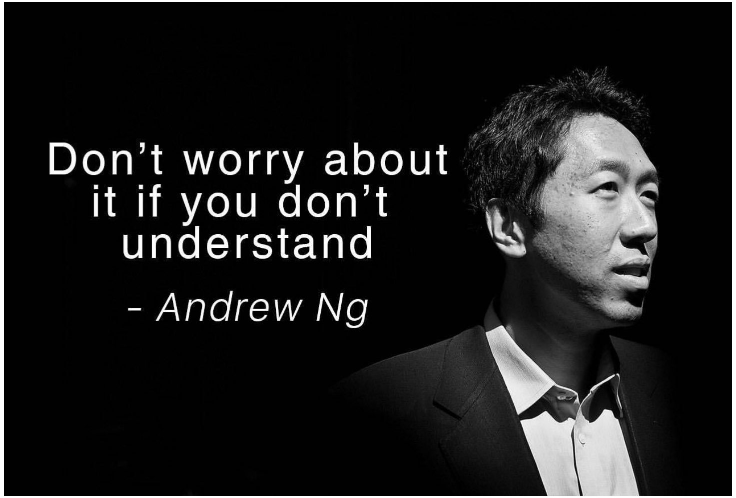
- How much do we want the TNG galaxies to overlap with real galaxies/how close should we allow them to be?
- How does masking in one band affect masking in others, since the galaxies will be different sizes in different bands?

Neural networks learn by updating weights iteratively according to some loss function; they define their own features

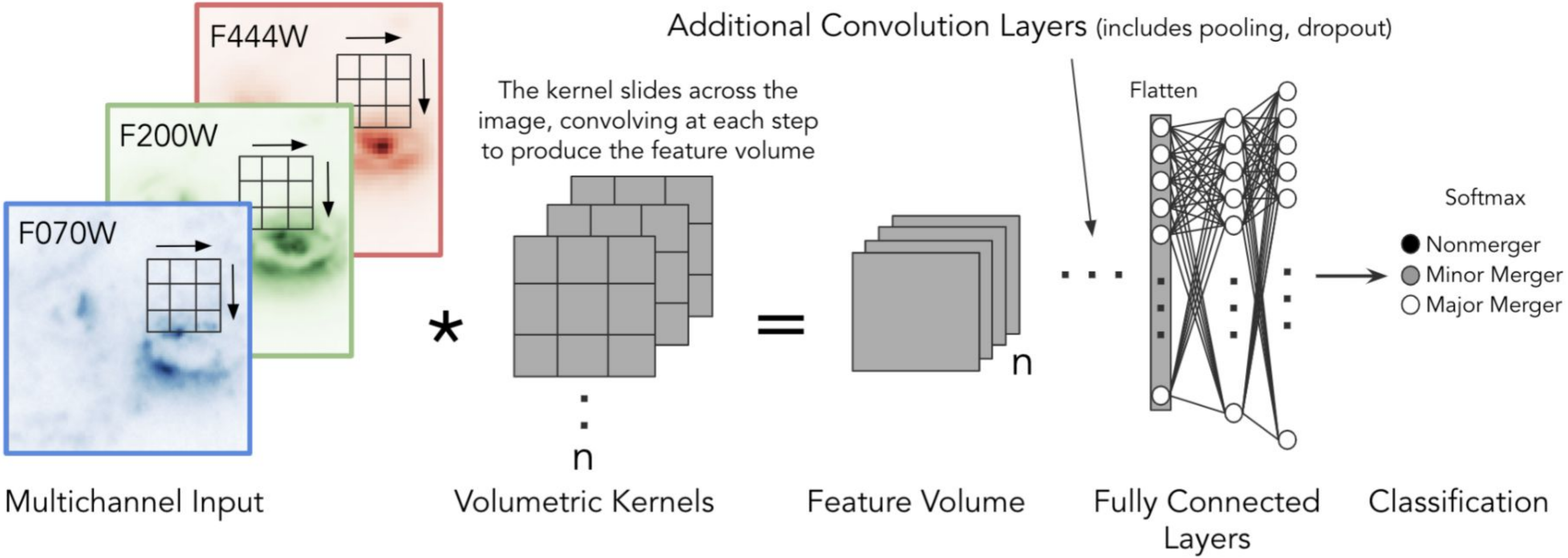




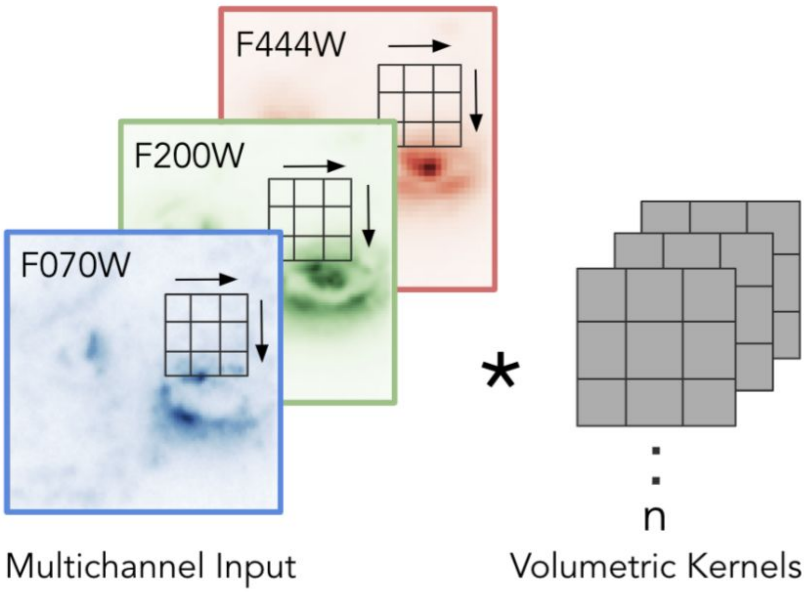
Resources for learning about neural networks and CNNs:
3Blue1Brown
Andrew Ng's Coursera course (also on youtube)



Convolutional Neural Networks have layers upon layers of convolution filters that extract features

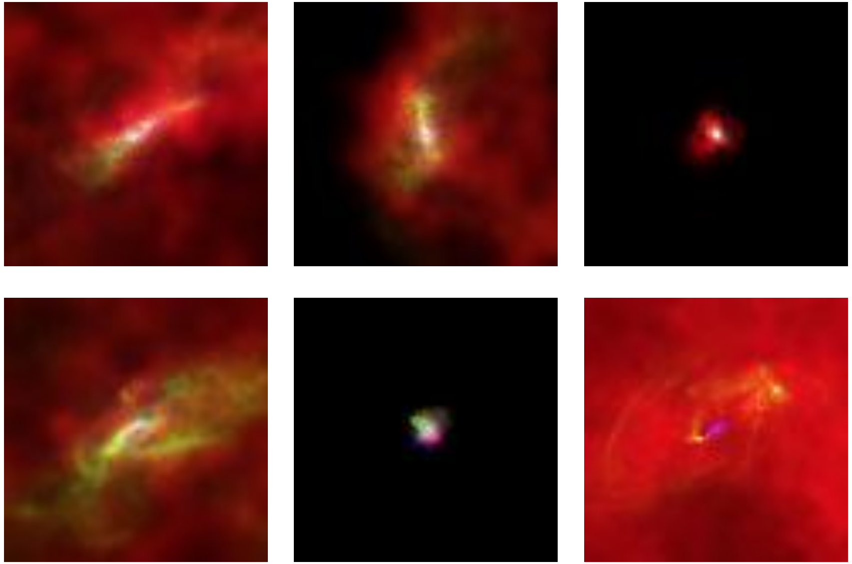


CNNs are optimal for multi-band image classification



- They learn filters in parallel
- Flexible
- Use multi-band input and deal with features from different bands in a spatially coherent way
- Relatively agnostic to location in image of feature

Aimee trained an AlexNet-esque CNN to identify merging and non-merging galaxies at z=0.2 and 1



Red = metals
Green = gas
Purpleish = stars

Discuss:

Which filters do you think will be the best for identifying mergers? (we can take bets now and then see which ones the network chooses later!)

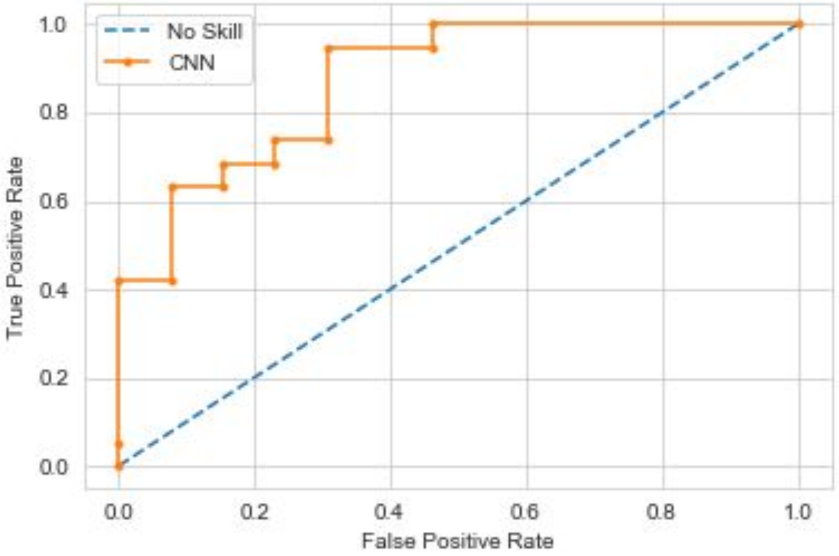
OR

Which wavelengths do you think are most important, since filters will show different features at different redshifts?

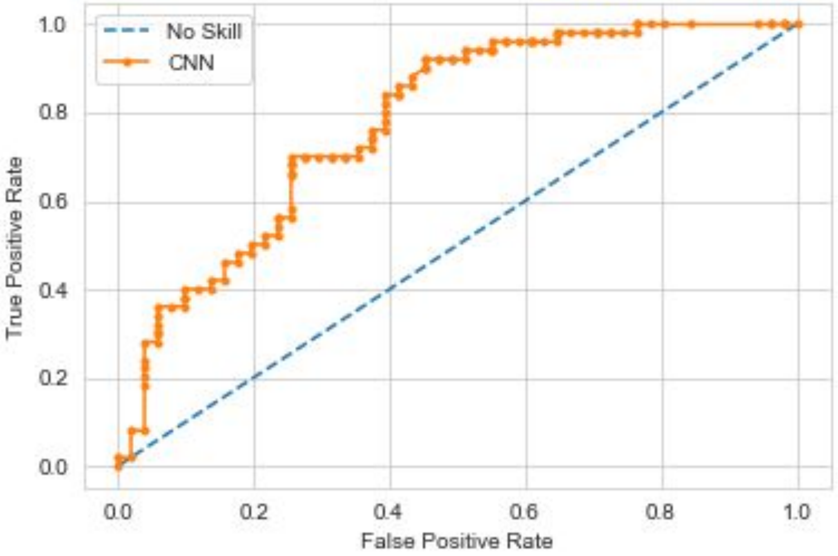
ROC curves show that the network learned!

The area under the curve is better than 0.5 (random guessing)

z = 0.2

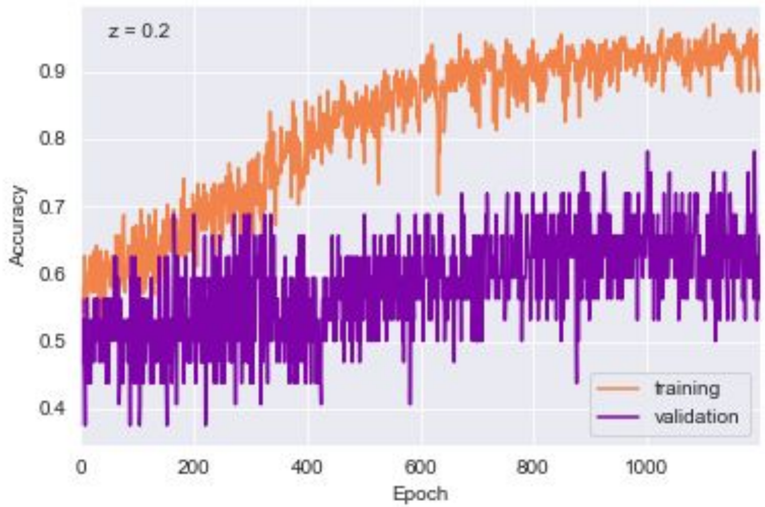


z = 1

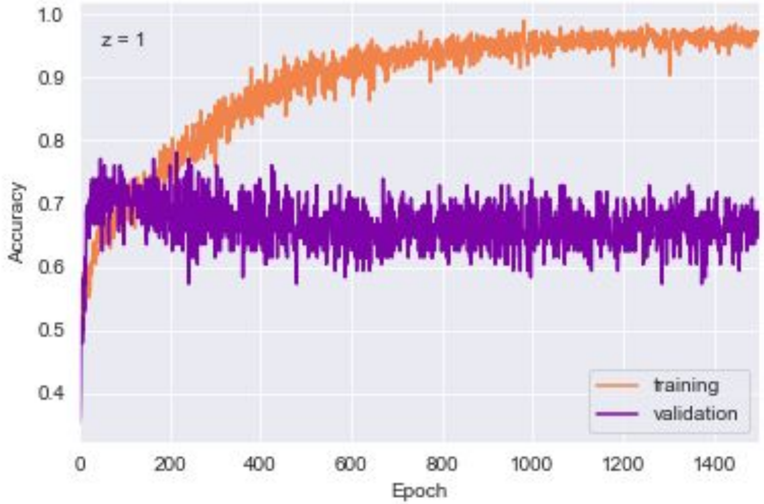


Accuracy Curves show that the CNN makes the right prediction about 65% of the time

z = 0.2

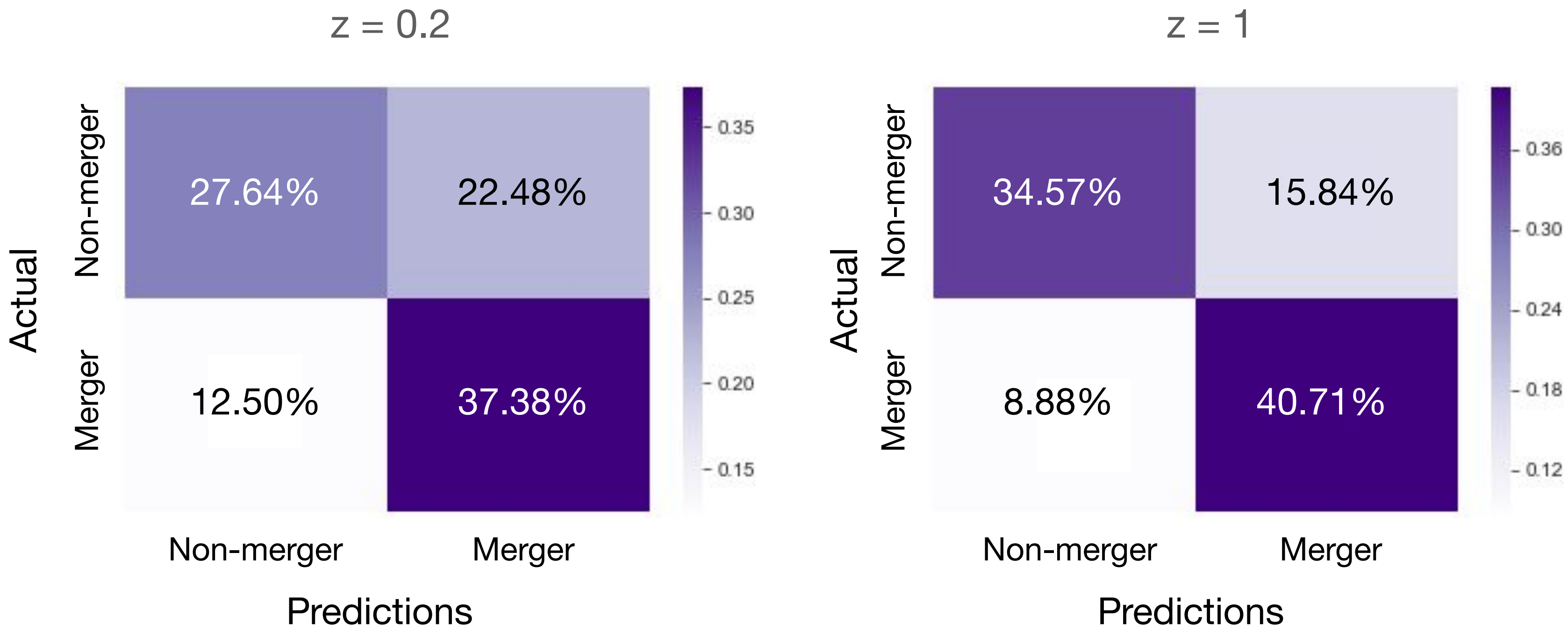


z = 1



We want to make sure we're not missing any mergers

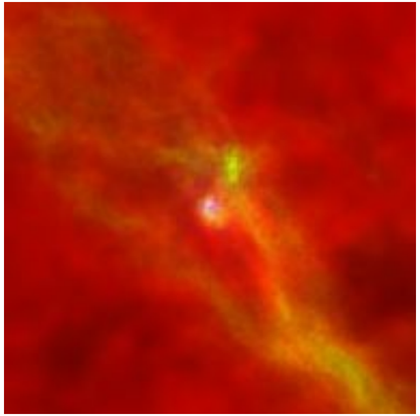
False positives are better than false negatives



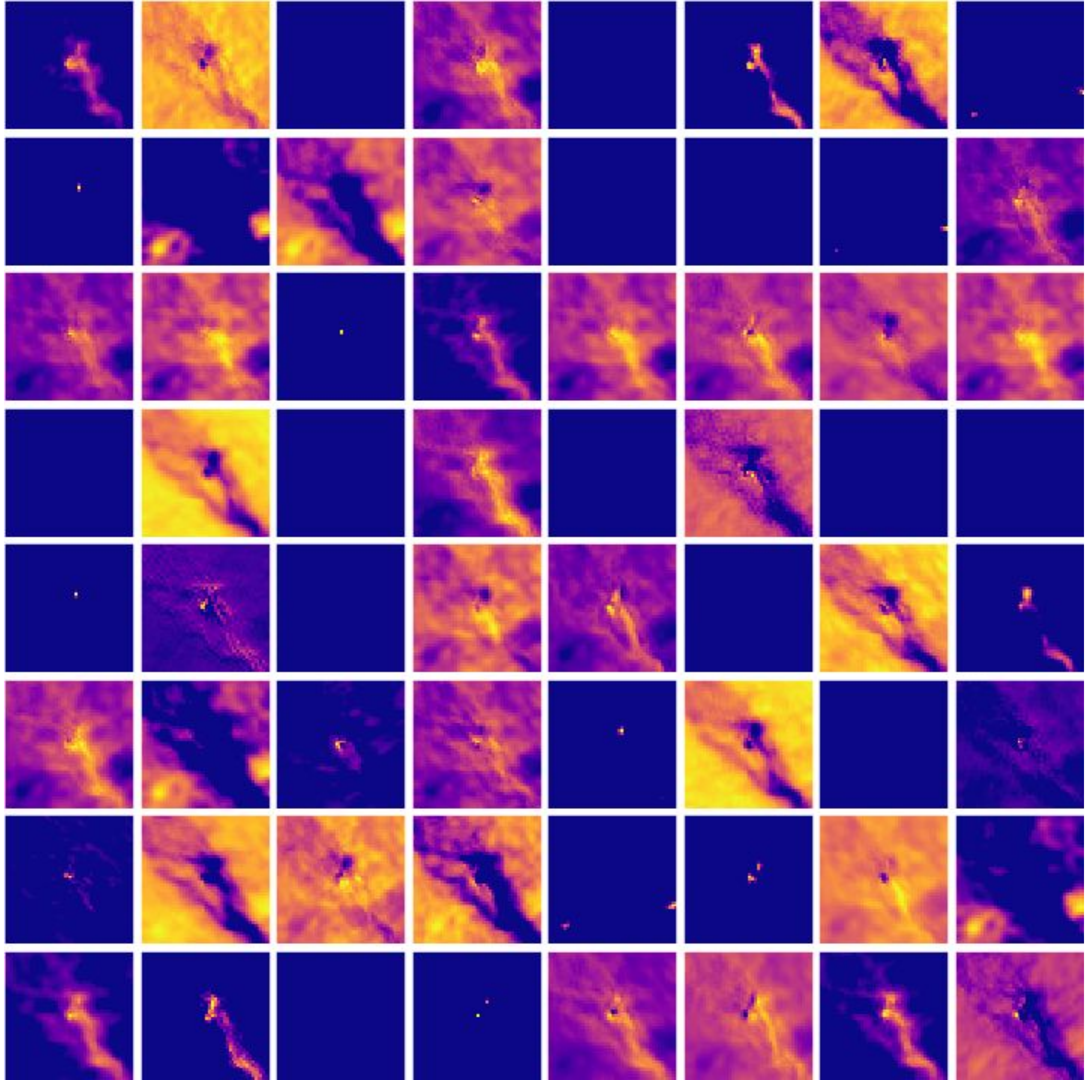
CNNs are interpretable!

Q1: What is the network actually looking at in its convolutional layers?

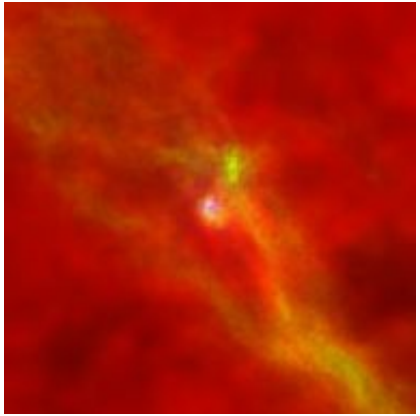
Merger at z = 0.2



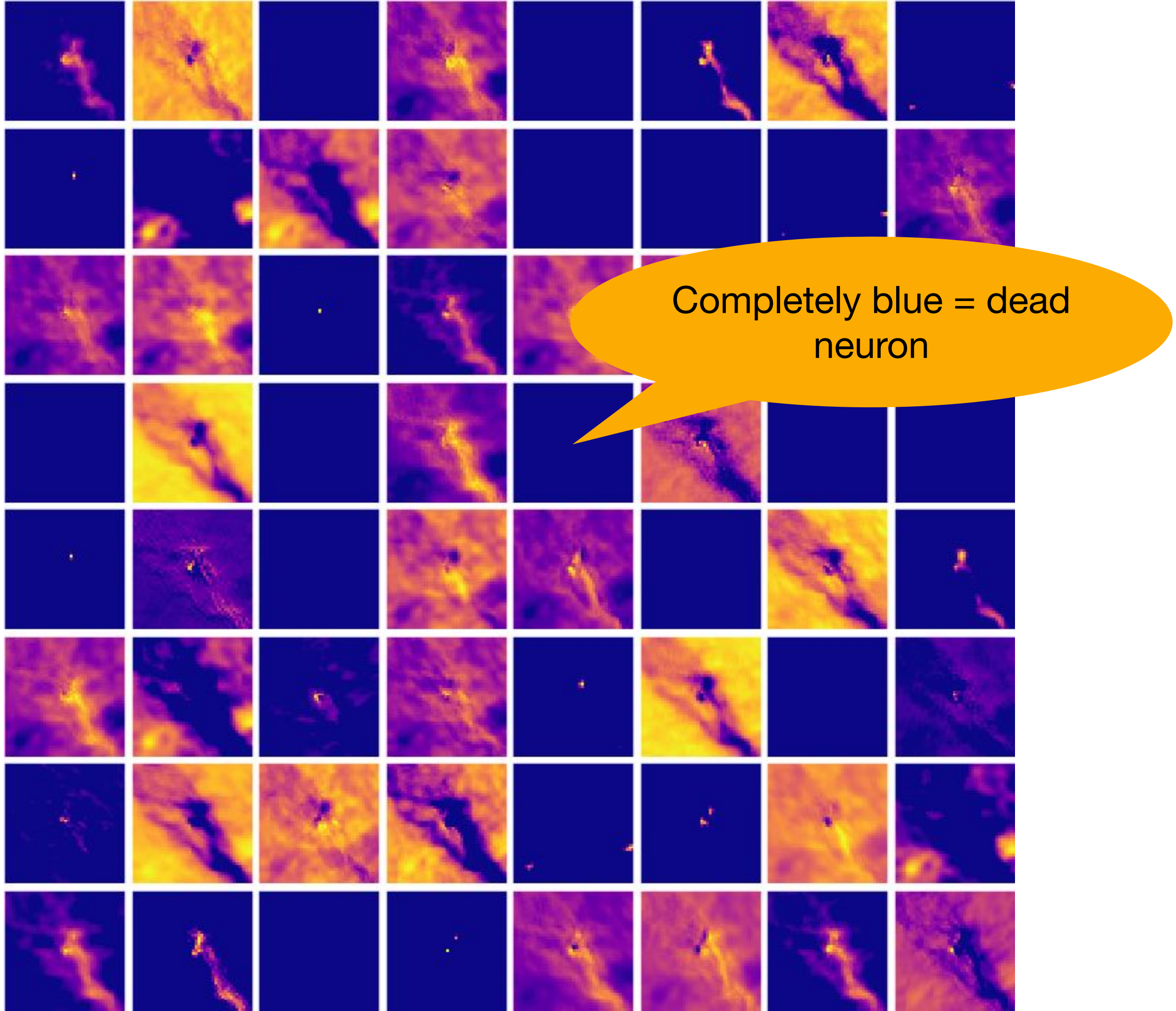
These filter activations on the left still look somewhat like the galaxy above…



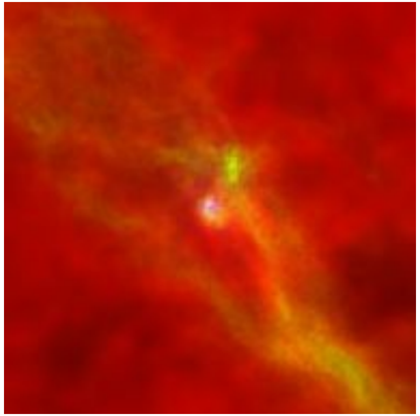
Merger at z = 0.2



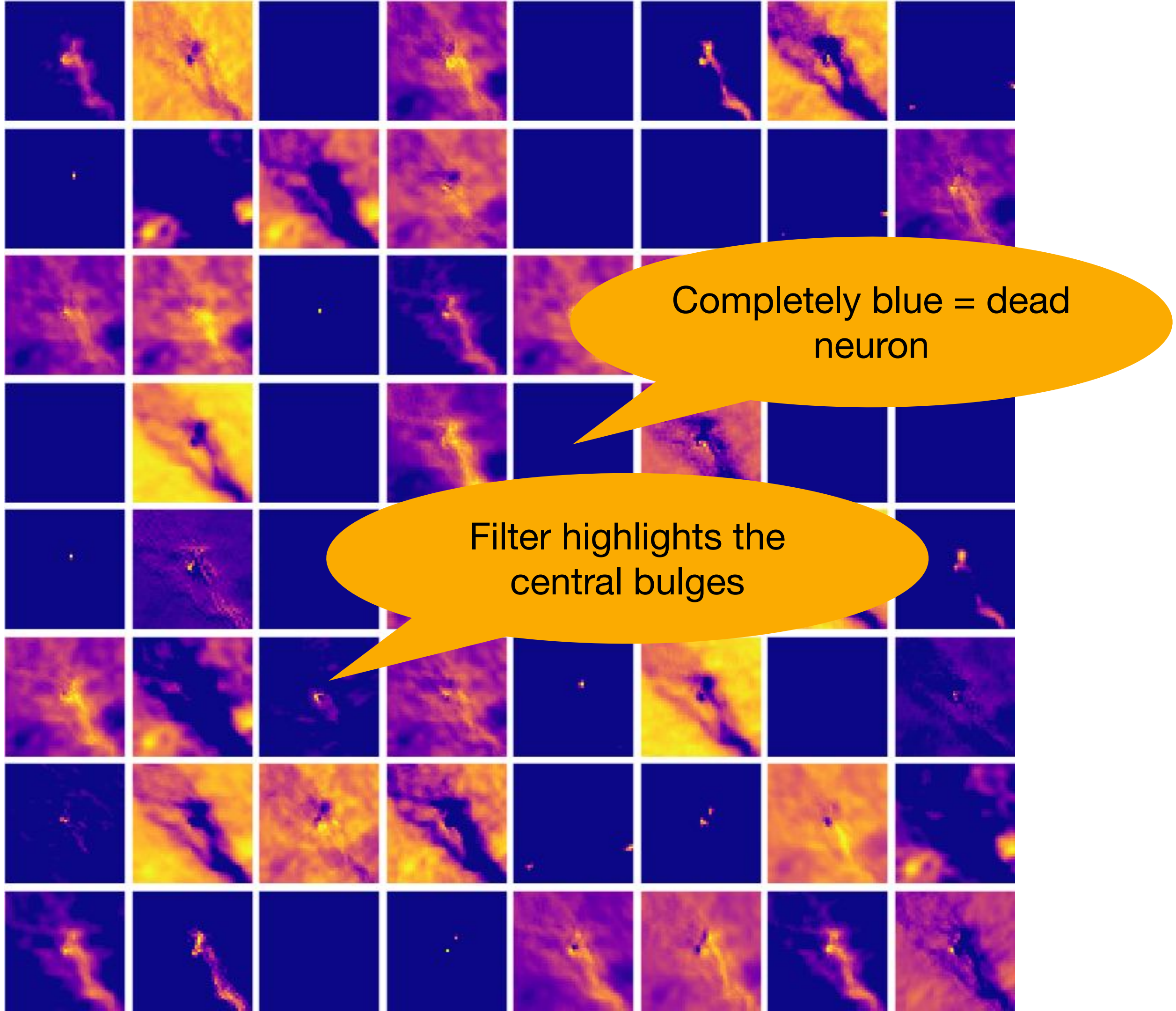
These filter activations on the left still look somewhat like the galaxy above…



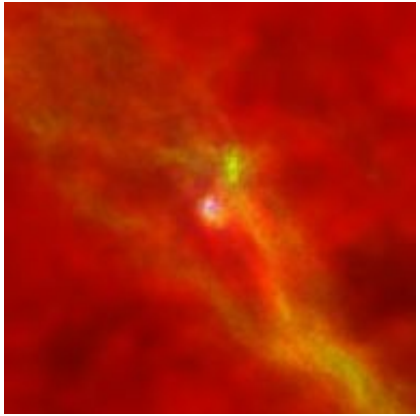
Merger at z = 0.2



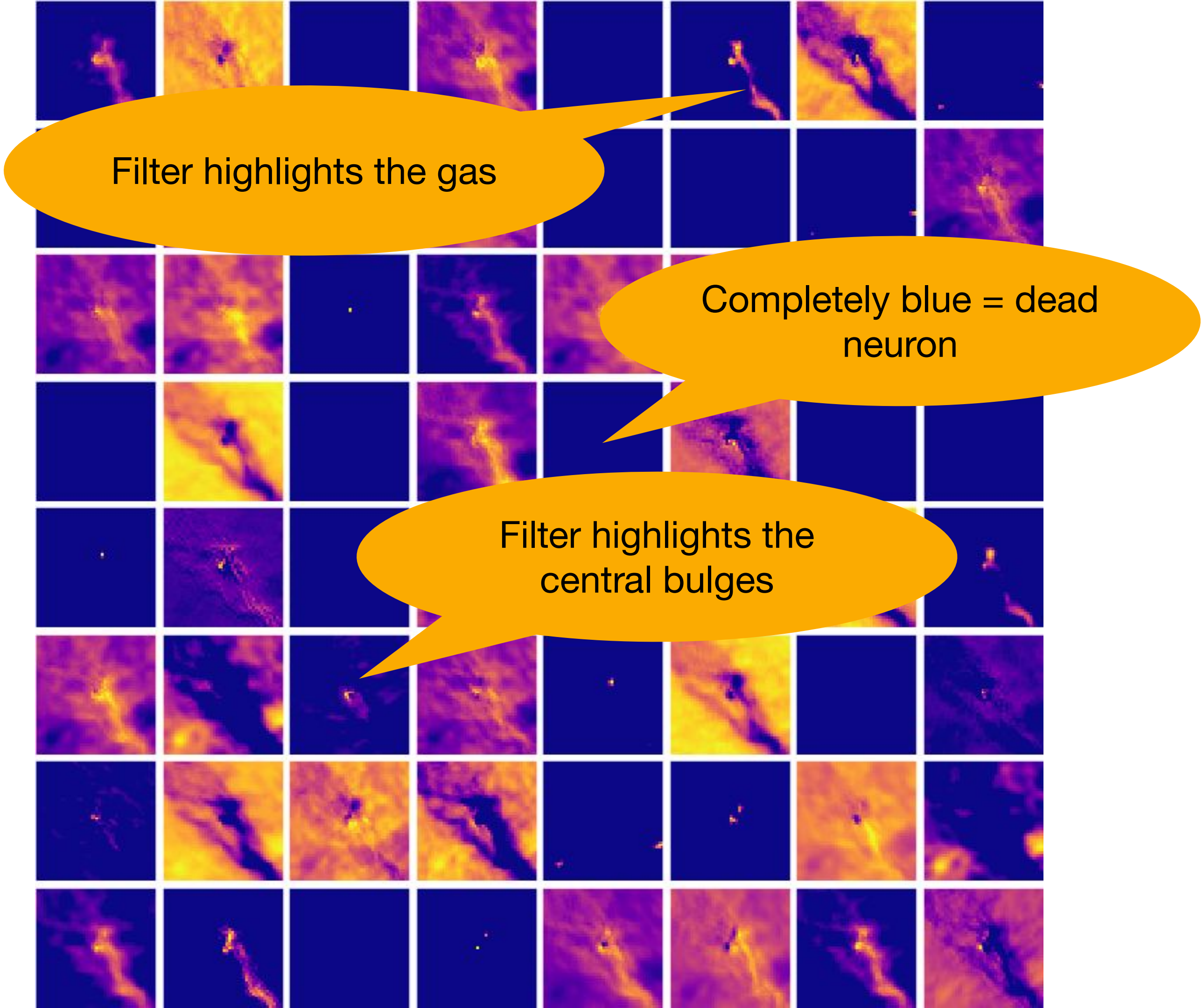
These filter activations on the left still look somewhat like the galaxy above…



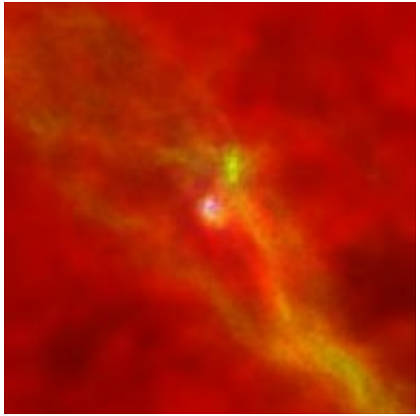
Merger at z = 0.2

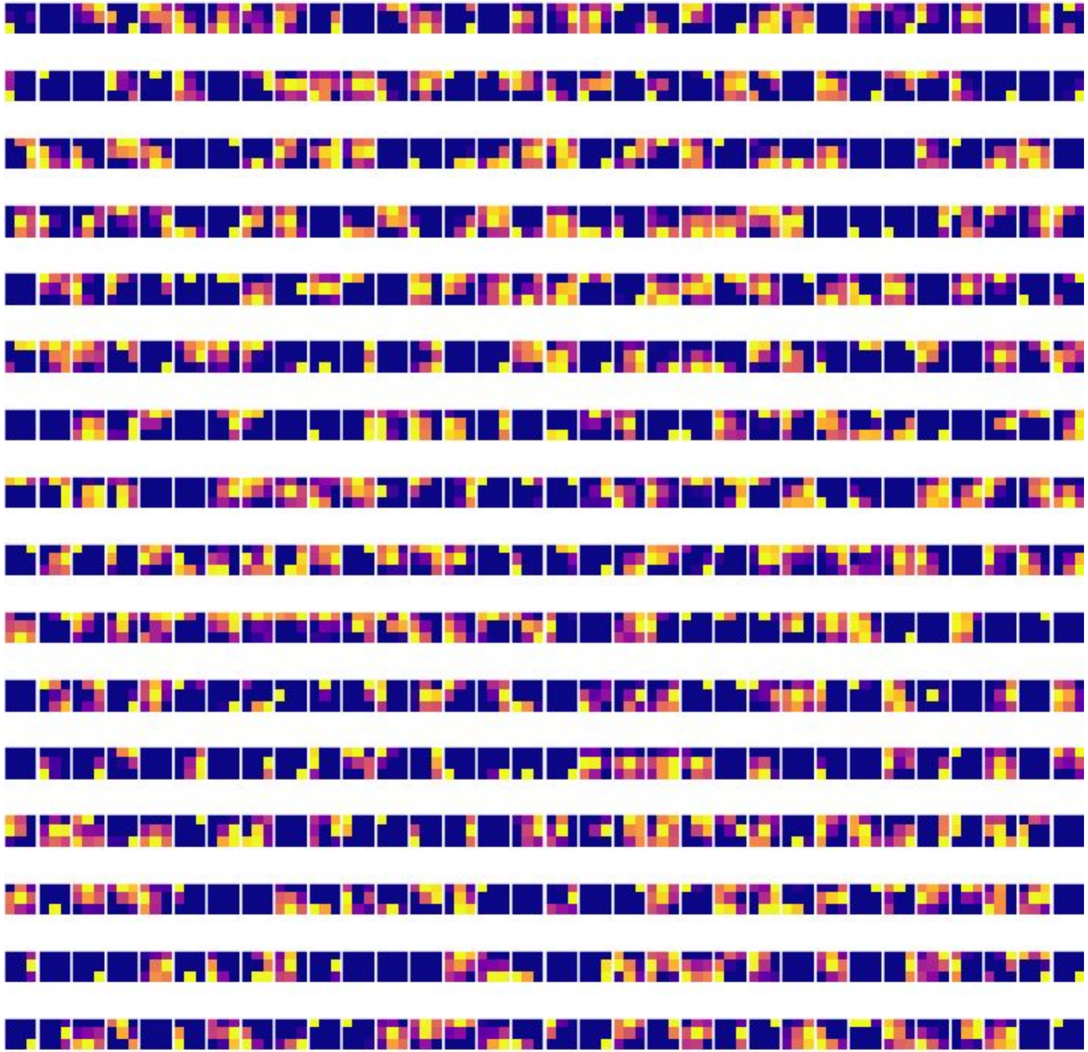


These filter activations on the left still look somewhat like the galaxy above…



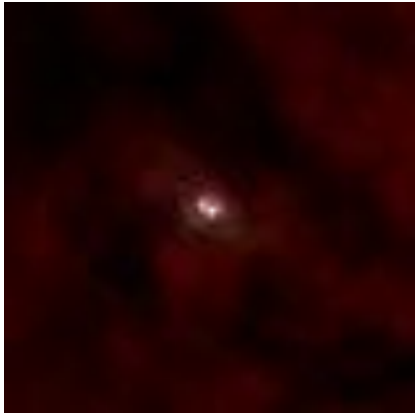
Merger at $z = 0.2$

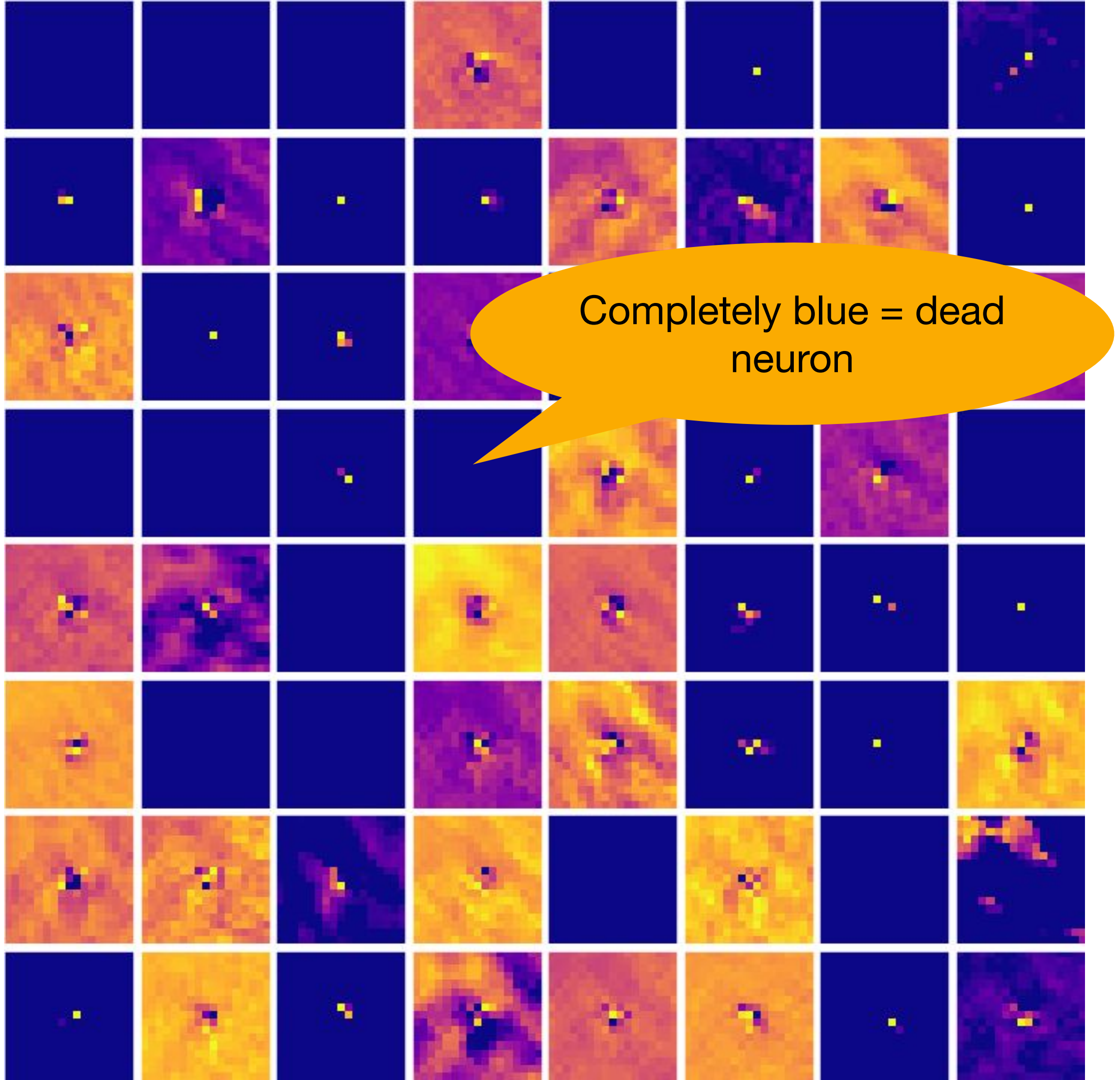




These filter activations don't look anything like galaxies anymore!

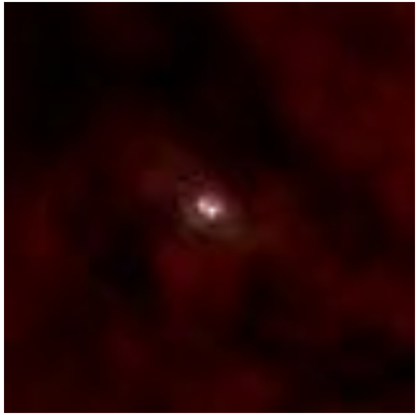
Merger at z = 1



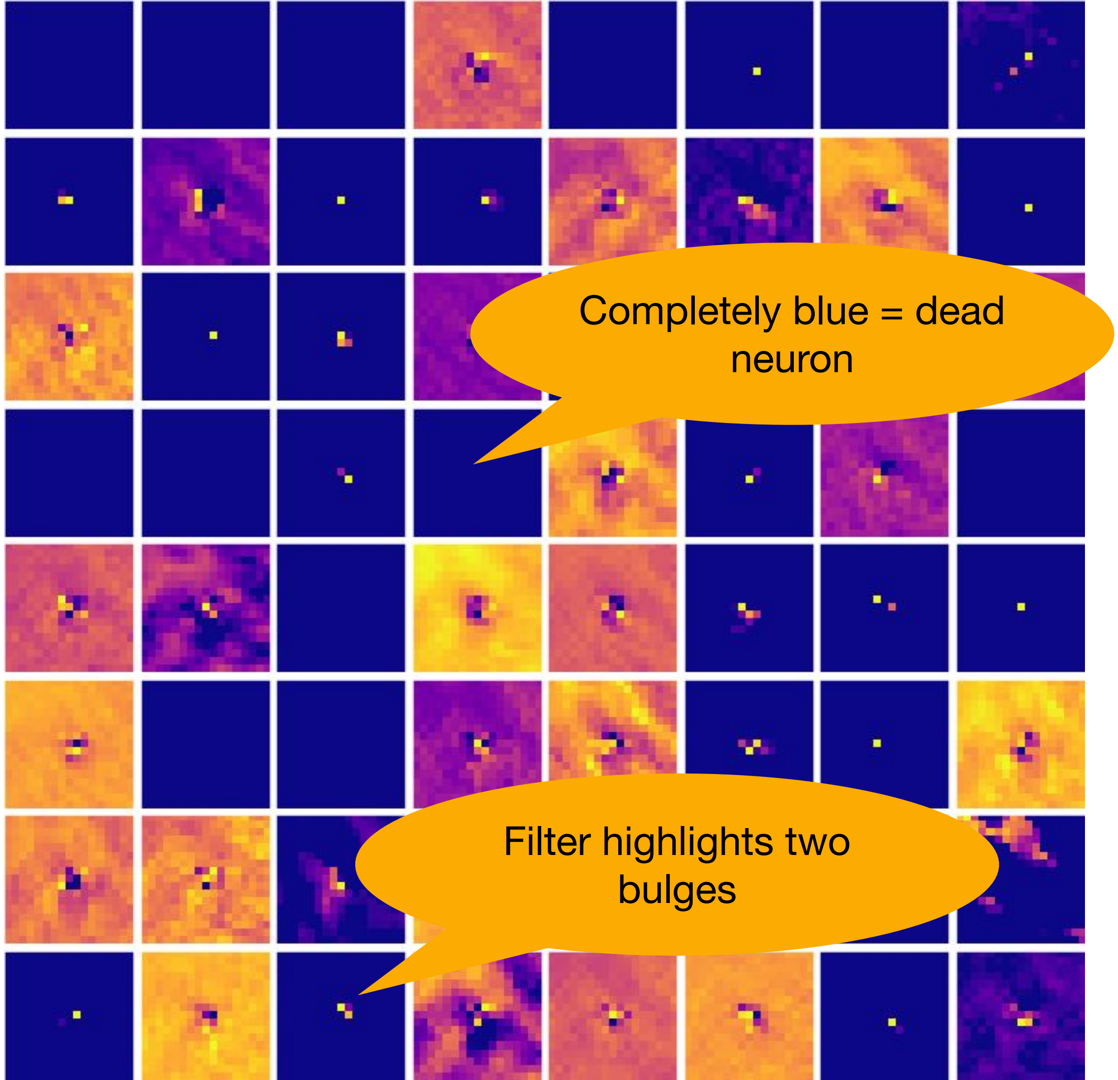


These filter activations look somewhat like galaxies…

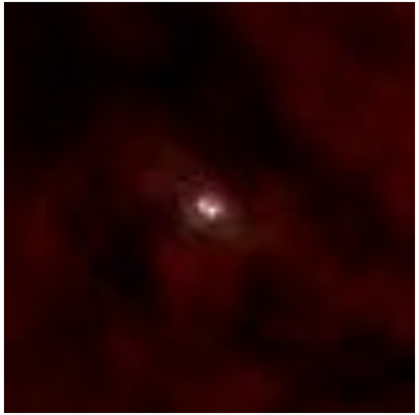
Merger at z = 1



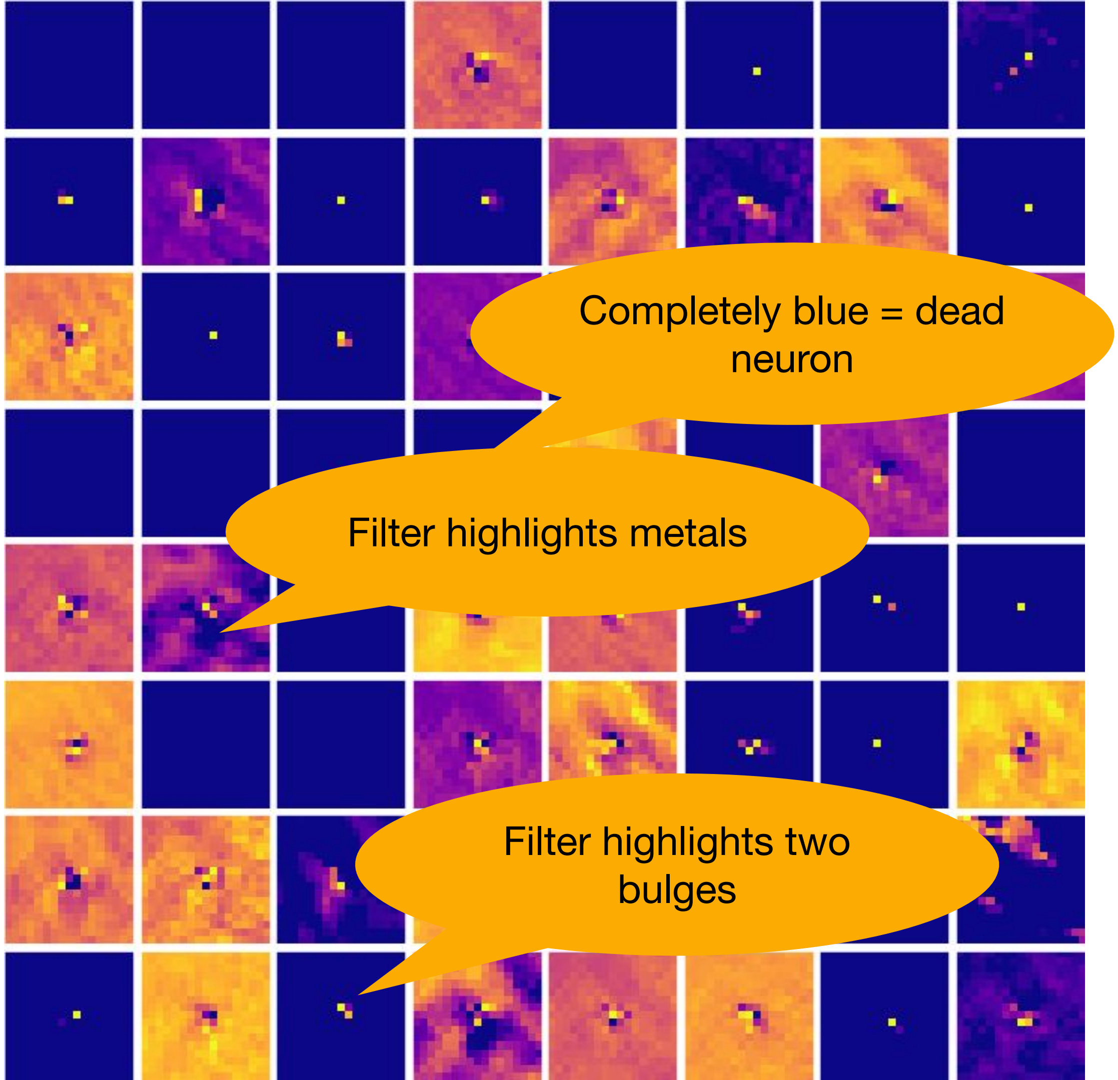
These filter activations look somewhat like galaxies…



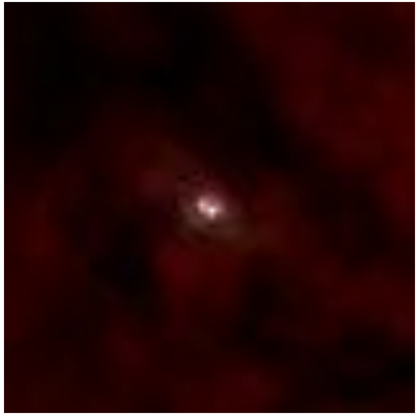
Merger at z = 1

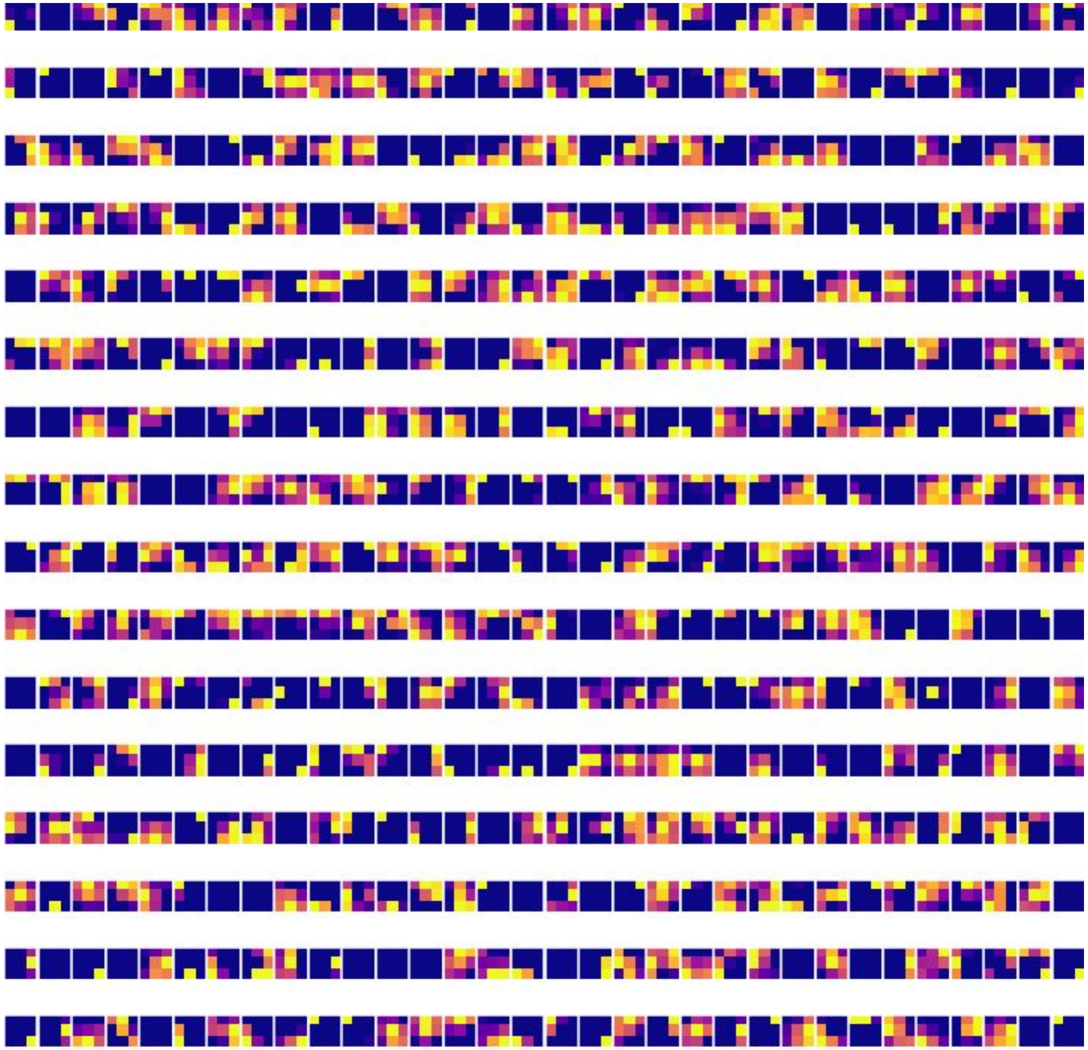


These filter activations look somewhat like galaxies…



Merger at z = 1



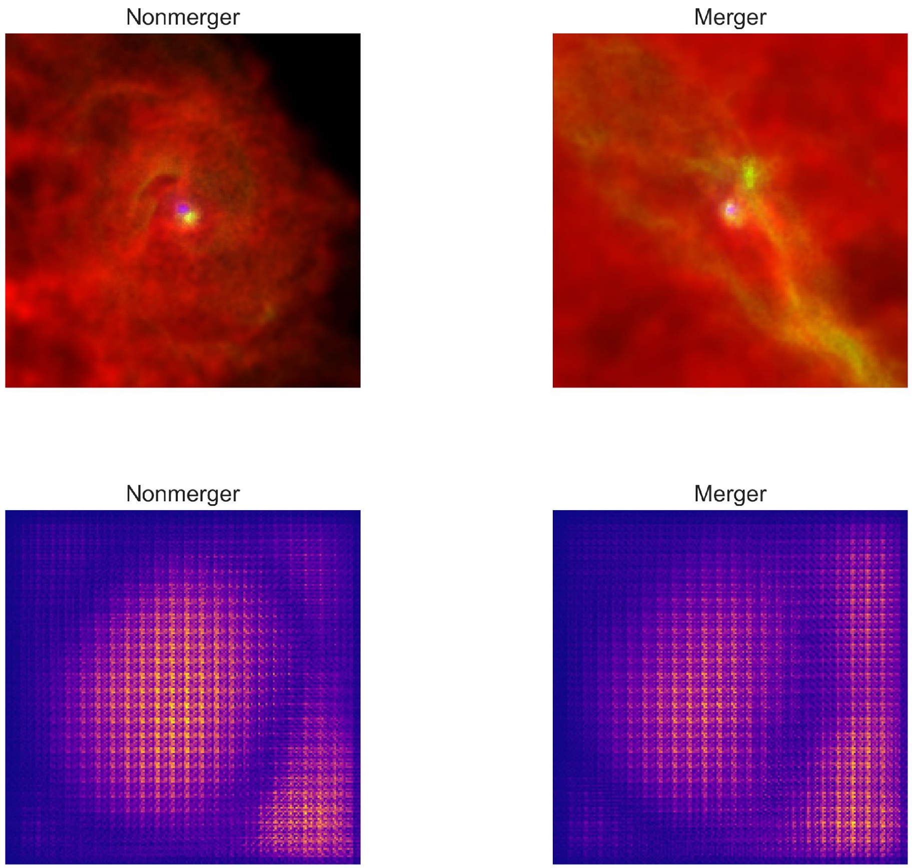


These filter activations again look nothing like galaxies!

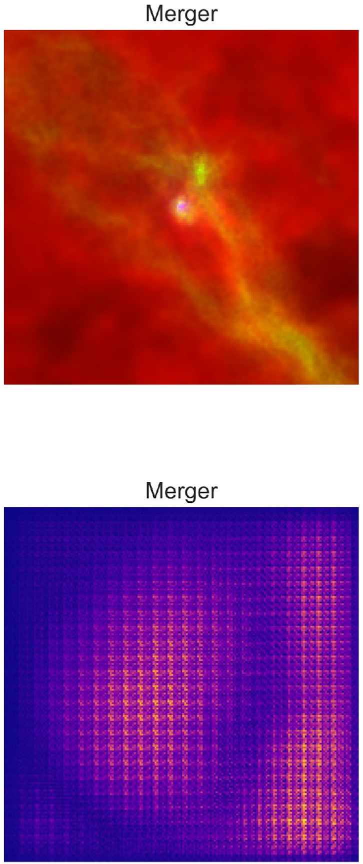
Q2: Where in the image is the CNN focusing to make a classification?

Merger at $z = 0.2$

Saliency maps measure how important each pixel is to the final classification. The brighter the pixel, the more important it is.

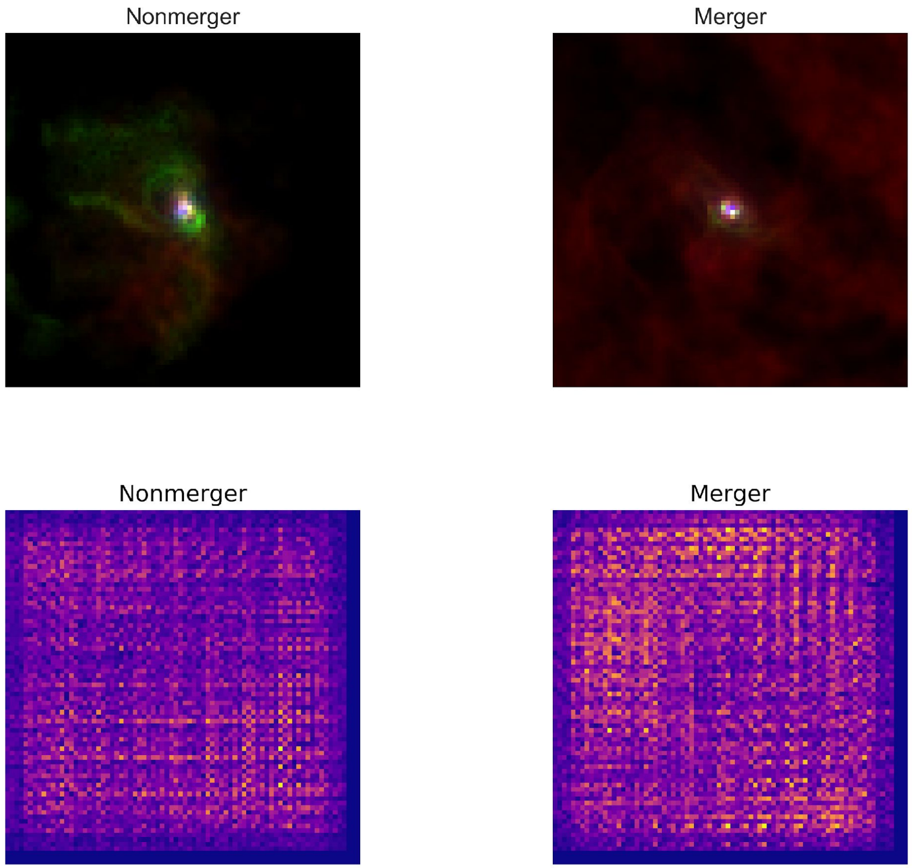


Merger at $z = 0.2$



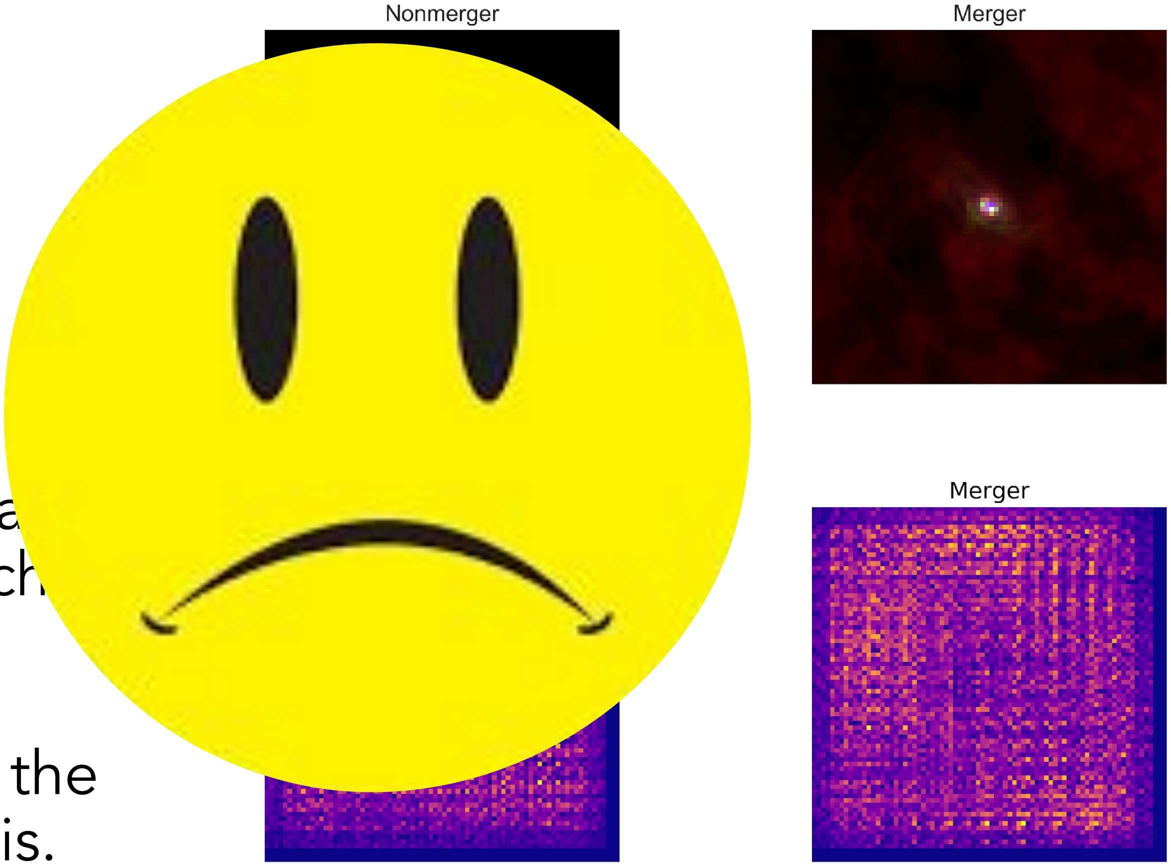
Saliency maps me
how important eac
pixel is to the final
classification. The
brighter the pixel, the
more important it is.

Merger at $z = 1$



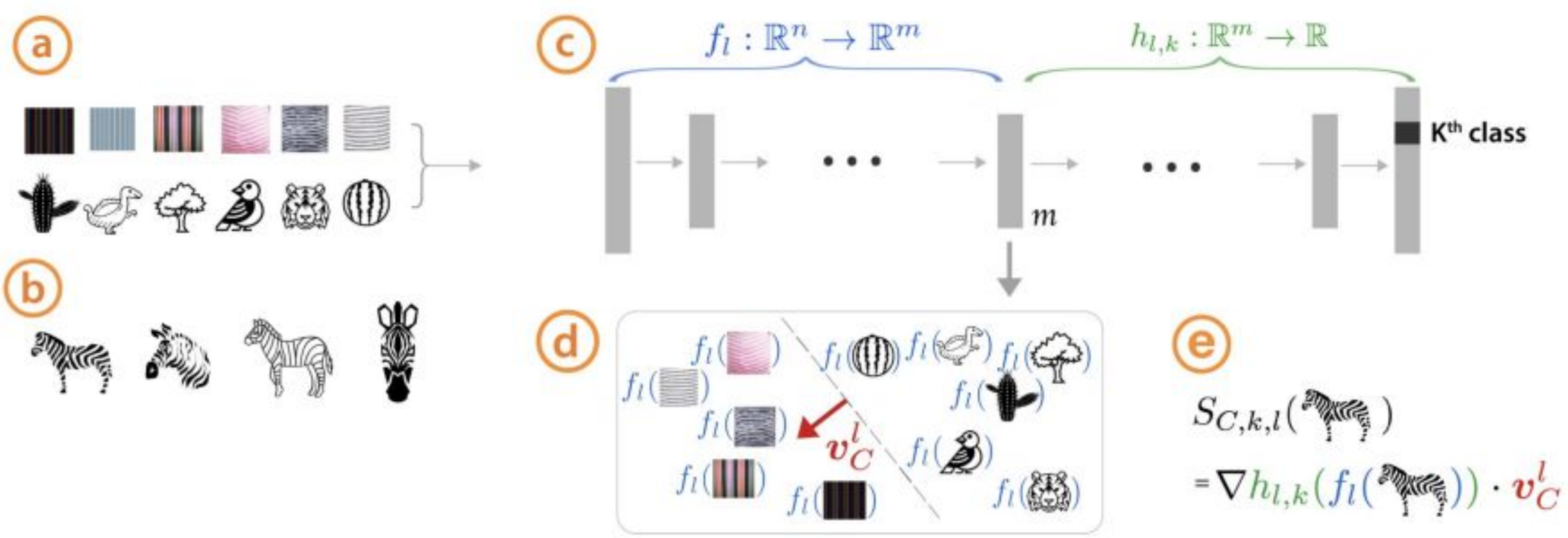
Saliency maps measure how important each pixel is to the final classification. The brighter the pixel, the more important it is.

Merger at $z = 1$



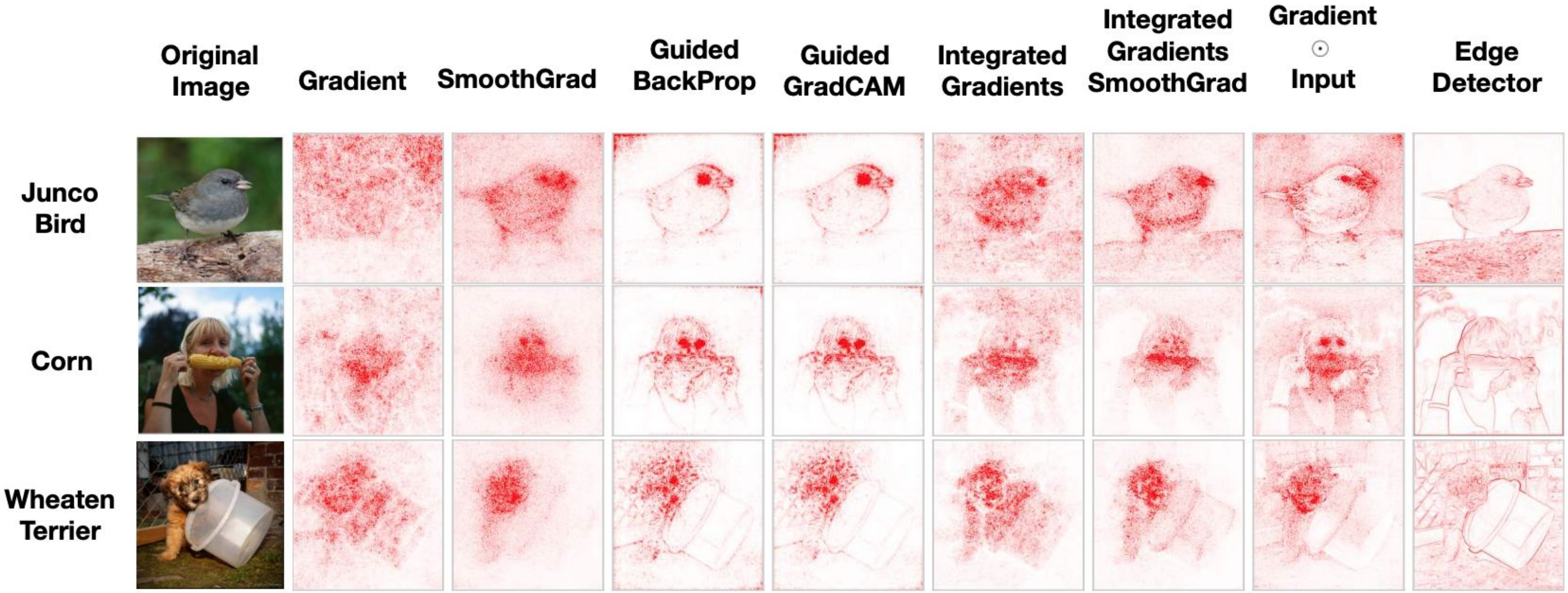
Saliency maps mea
how important each
pixel is to the final
classification. The
brighter the pixel, the
more important it is.

TCAVs: Testing with concept activation vectors allows humans to test if the network learns concepts



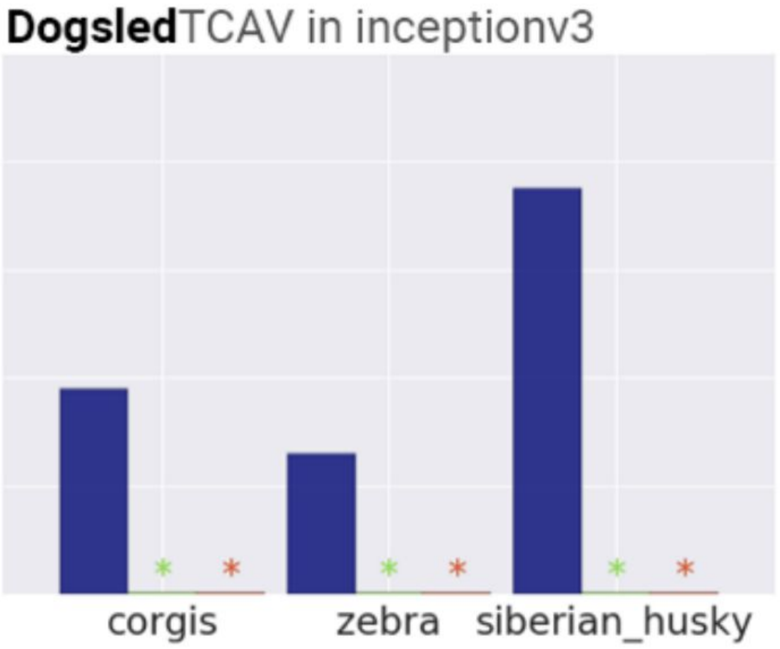
Interpretability beyond feature attribution: Kim+2018 https://arxiv.org/pdf/1711.11279.pdf,
also https://www.youtube.com/watch?v=Ff-Dx79QEEY&ab_channel=MLconf
ALSO Sanity Checks for Saliency Maps Adebayo+2018

Saliency maps can be a little sketchy



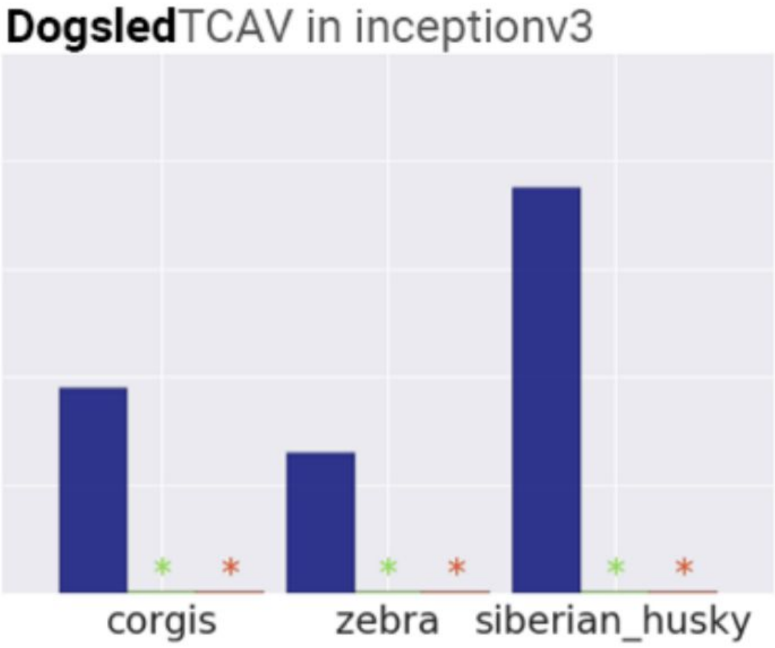
"Sanity Checks for Saliency Maps" Adebayo+2018

TCAVs: Testing with concept activation vectors offer global explanations for CNN decisions



Interpretability beyond feature attribution: Kim+2018 https://arxiv.org/pdf/1711.11279.pdf,
also https://www.youtube.com/watch?v=Ff-Dx79QEEY&ab_channel=MLconf

TCAVs: Testing with concept activation vectors offer global explanations for CNN decisions

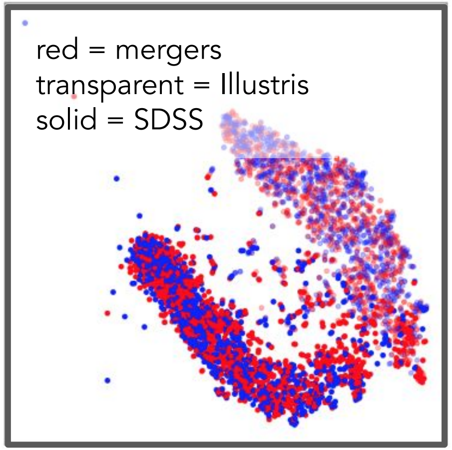


Ideas for galaxy-based CNNs:

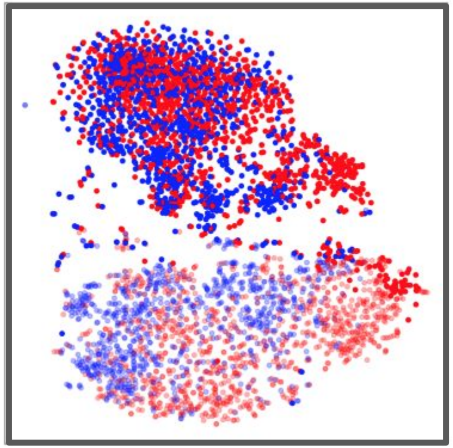
- ‘Gas-rich’ concept
- ‘Disky’ concept
- ‘Busy field’ concept

Domain adaptation finds invariant features between training and target data

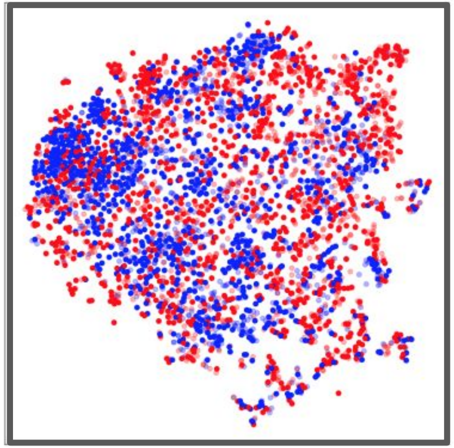
Before training the CNN



CNN, no domain adaptation

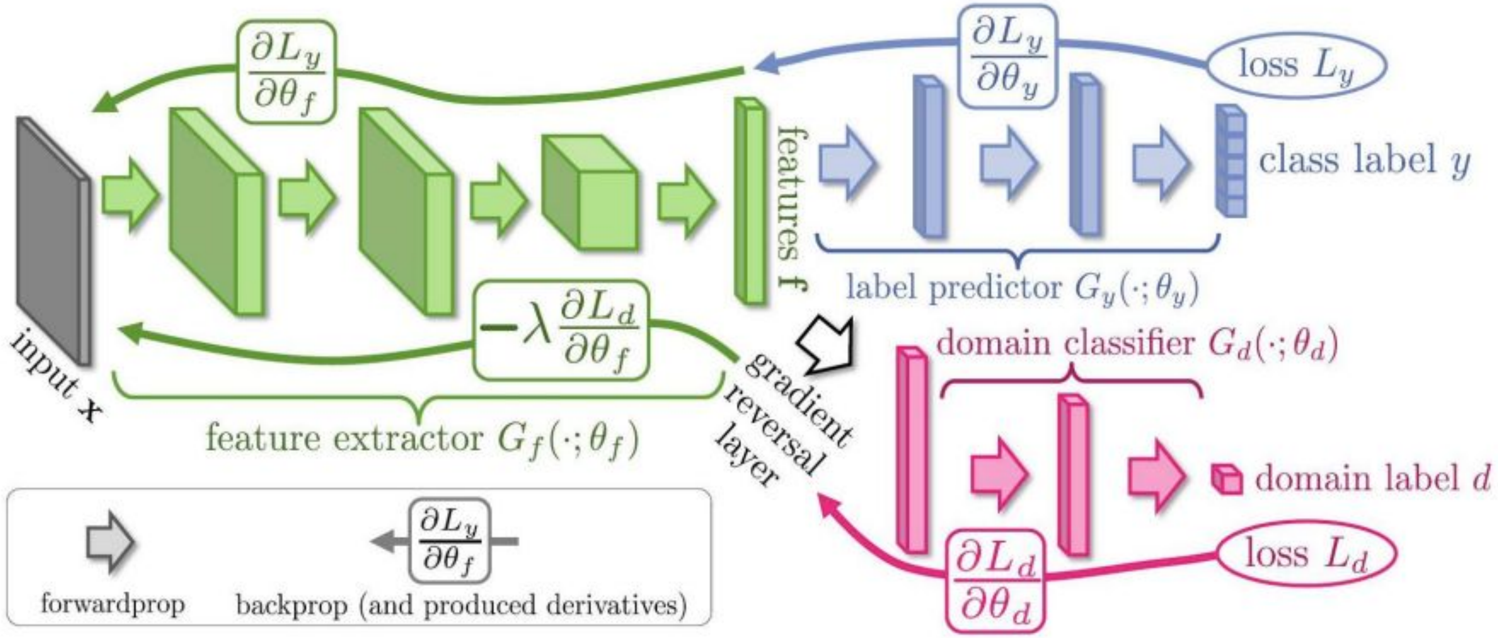


CNN and domain adaptation



t-SNEs from Alexandra Ciprijnovic's 2021 paper --^

Domain adaptation: The jump from TNG50 to *JWST* will require new architecture



Wang+2018

Discuss:
Domain adaptation will reveal differences between TNG and the real Universe? What would you be curious about?

Team ‘Fake it till you make it’

A smorgasbord of mocks from Illustris TNG50

HTST NIRCam

Becky Nevin

HST CANDELS

Aimee Schechter

SKIRT9 + AGN

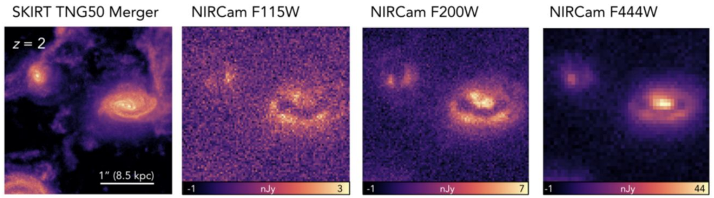
Jacob Shen

HSC-Joint,
MaNGA, SAMI, HECTOR

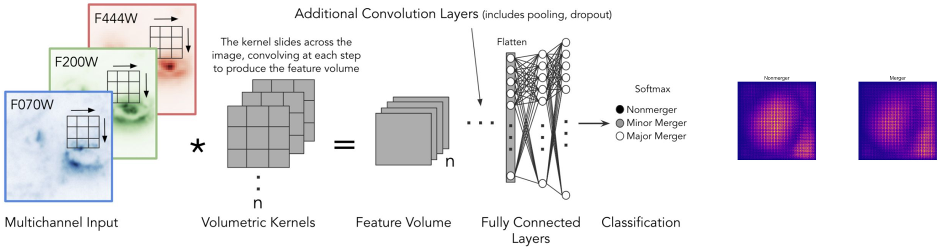
Connor Bottrell

Conclusions

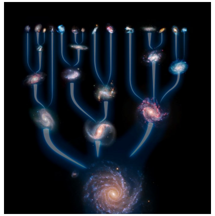
Realistic mock images are needed for accurate merger identification

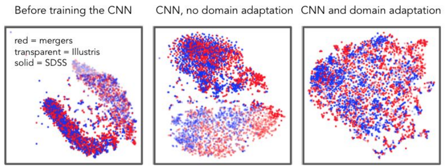


CNNs are an interpretable tool that can be used across redshifts and various merger stages



After identifying mergers from *HST* and *HTST* using domain adaptation, these merger catalogs can help us study the role of mergers in AGN, star formation, disk instabilities, and mass growth in the early universe

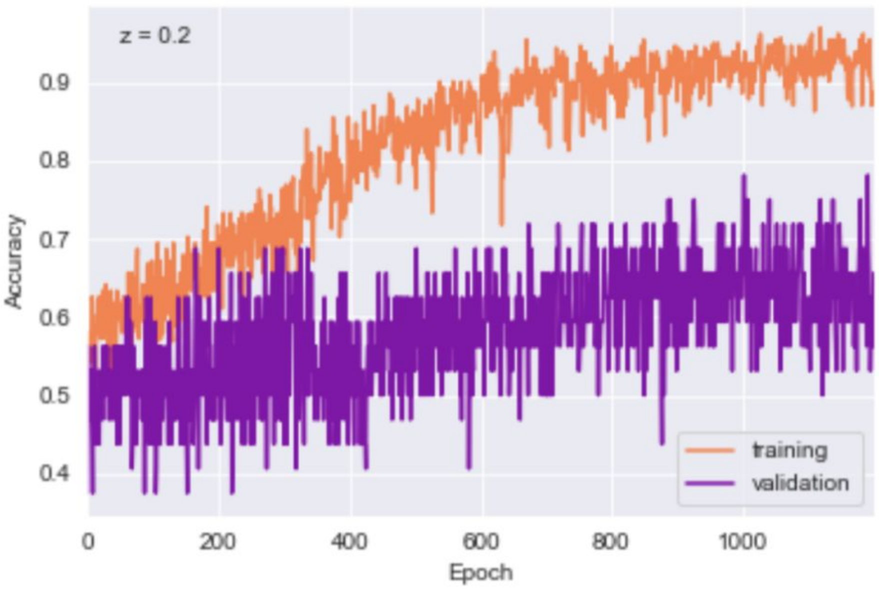




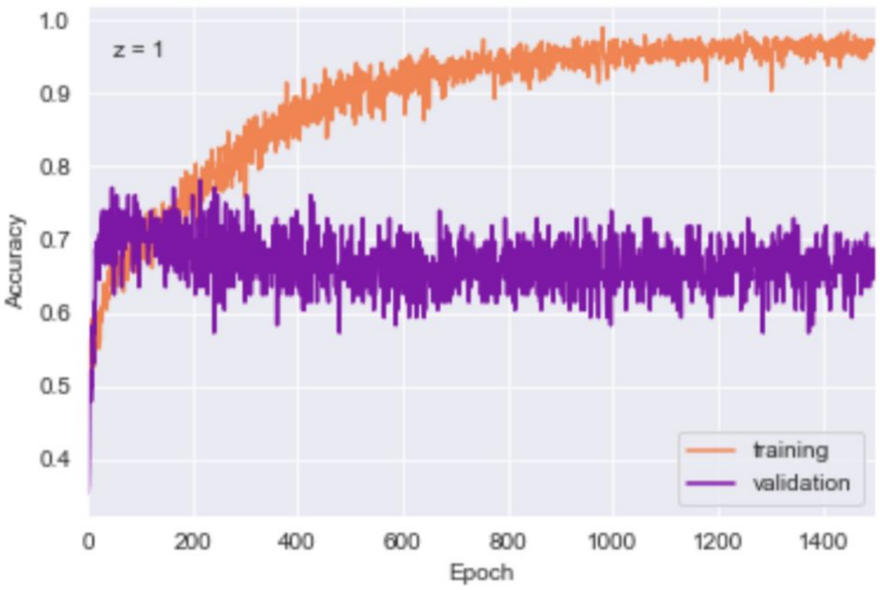
Conclusion slide

The learning curves show that the CNN makes the right prediction about 65% of the time

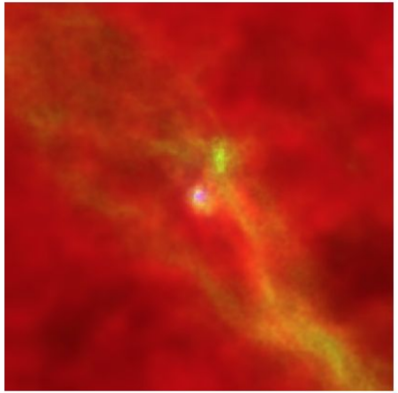
z = 0.2



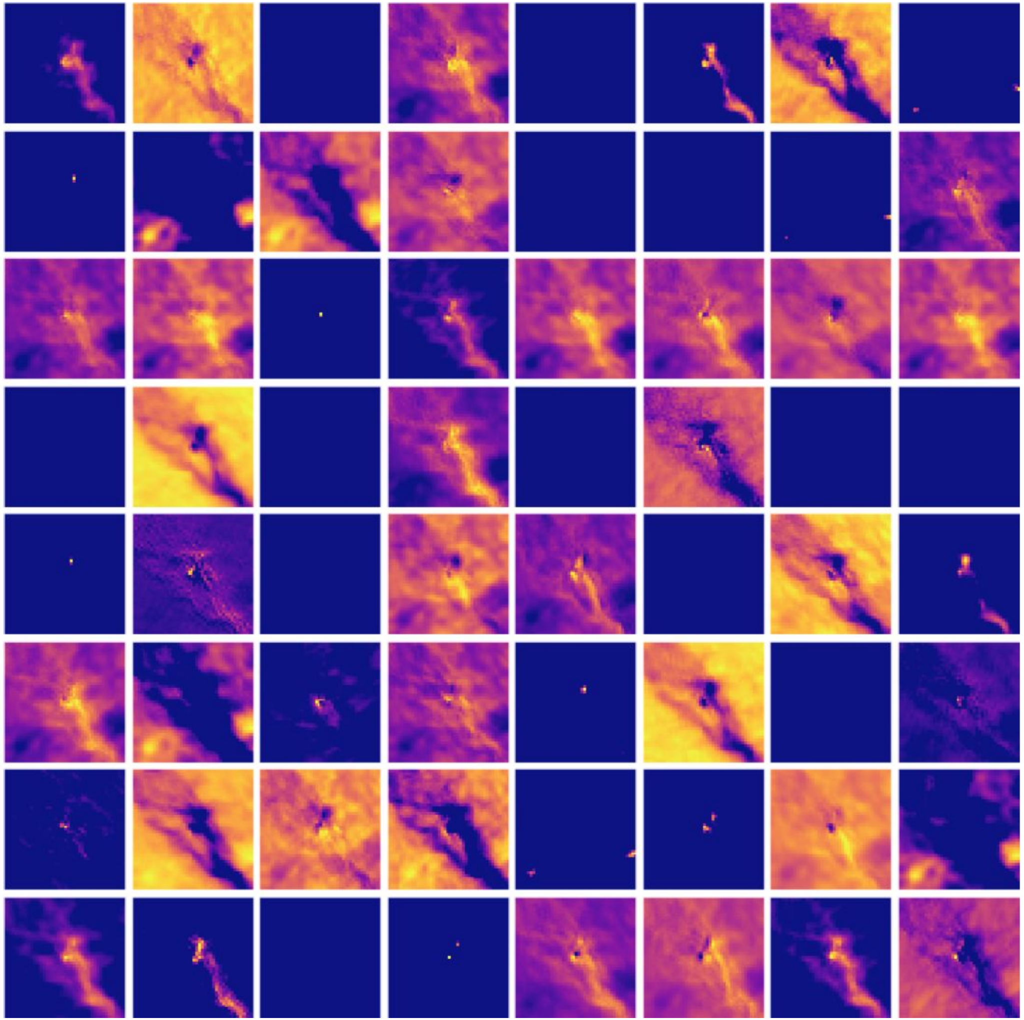
z = 1



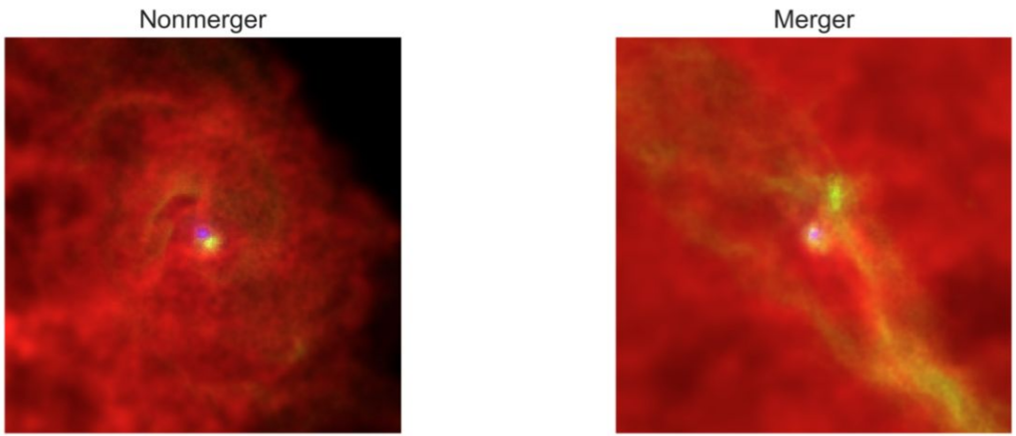
Merger at $z = 0.2$



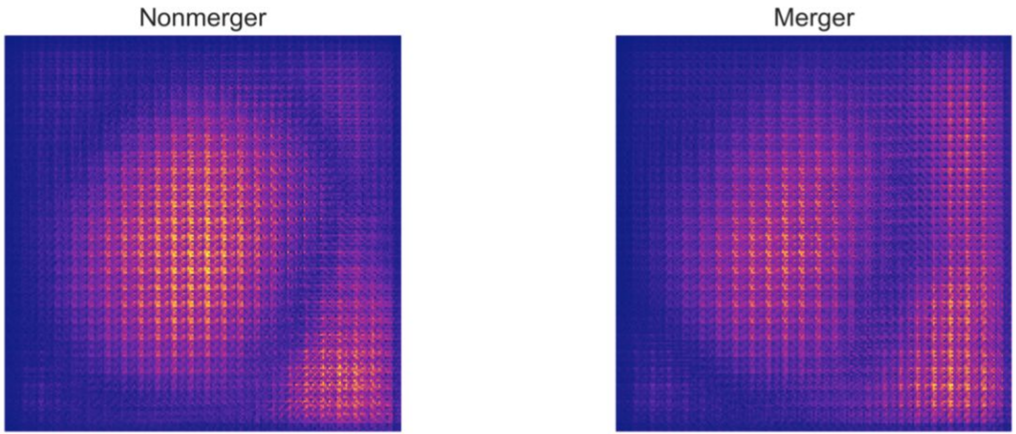
These filter activations on the left still look somewhat like the galaxy above…



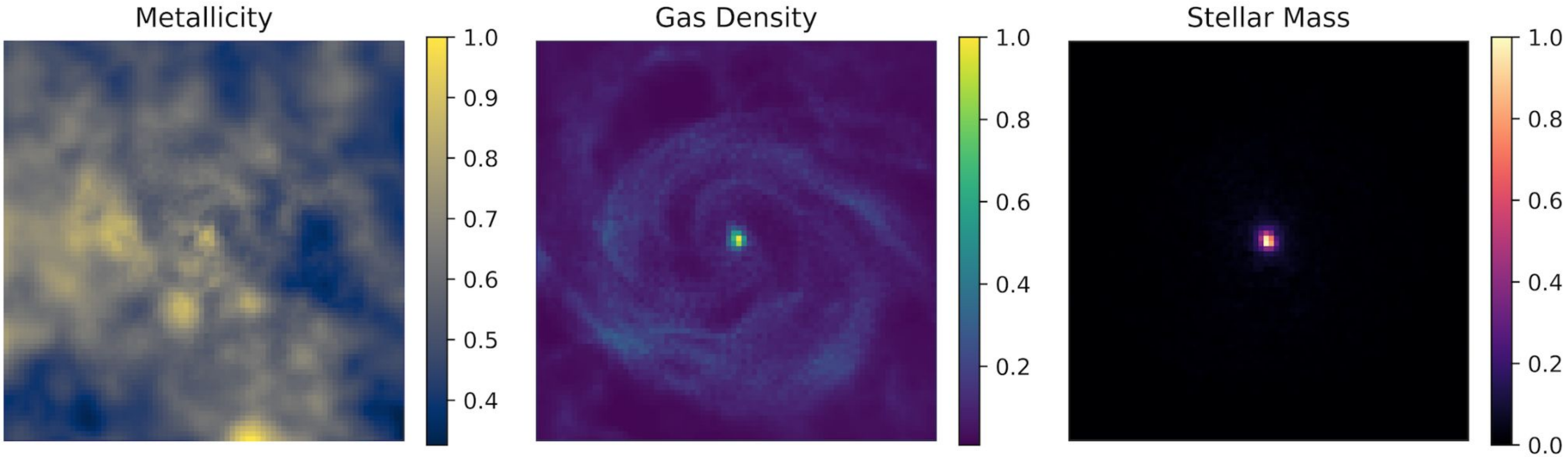
Merger at $z = 0.2$



Saliency maps measure how important each pixel is to the final classification. The brighter the pixel, the more important it is.

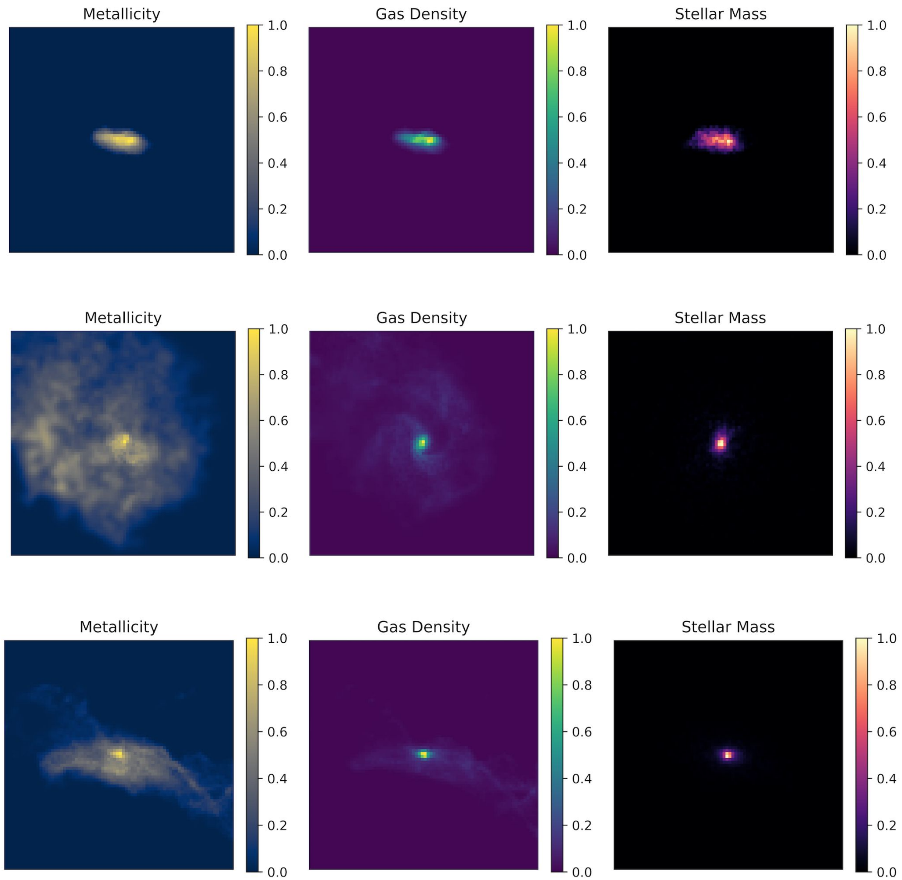


But, radiative transfer takes too long, so we use *yt* to create particle images

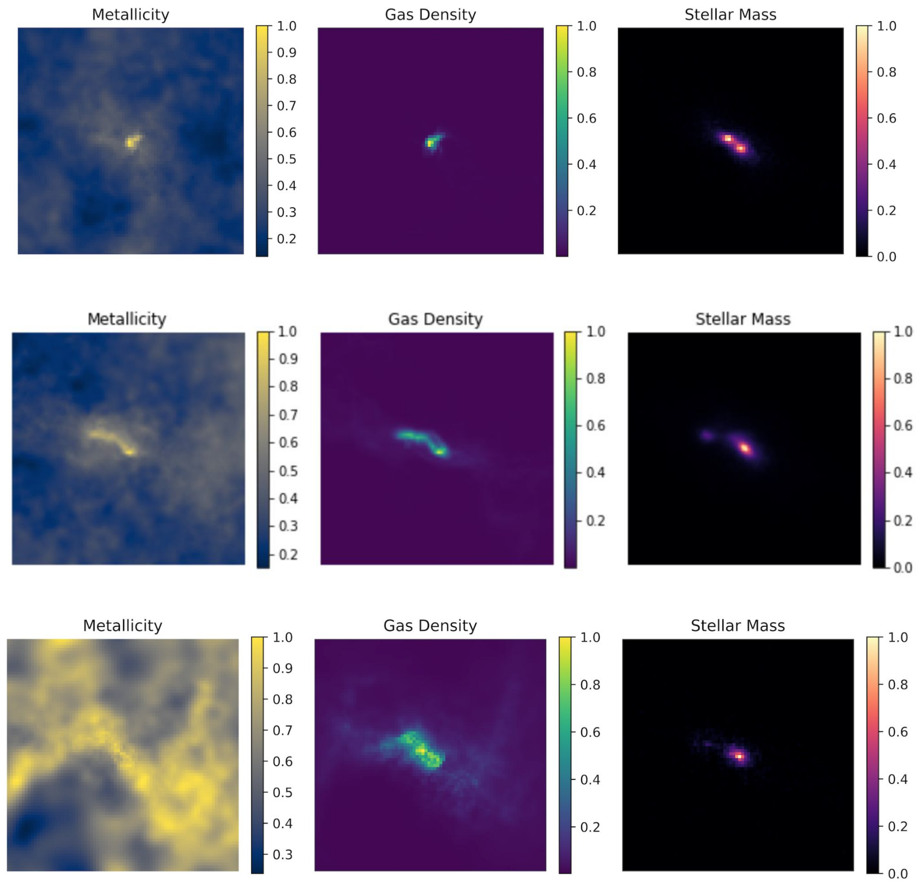


20 kpc width

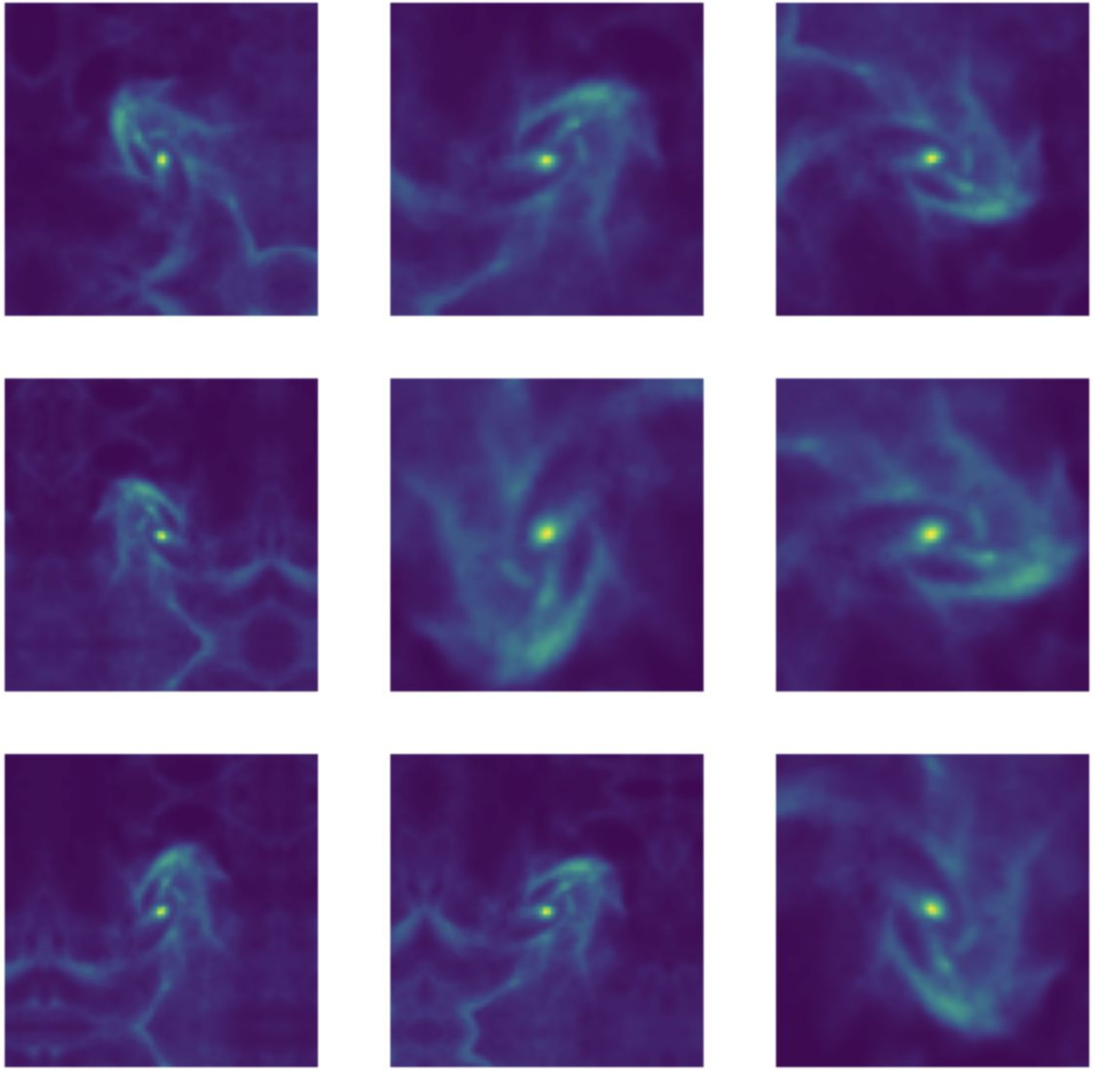
Non-mergers



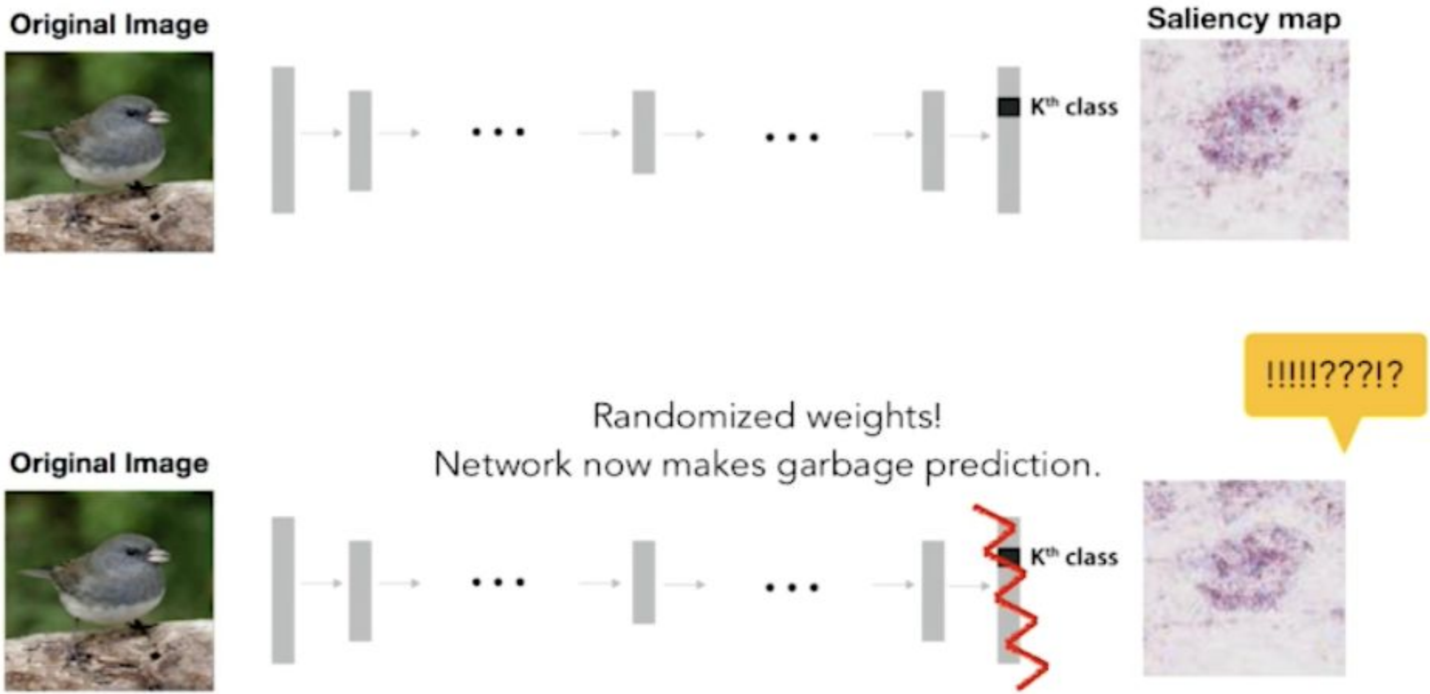
Mergers (pre, current, post)



Data augmentation adds to the sample size
- by how much?



Some confusing behaviors of saliency maps.



Sanity Checks for Saliency Maps
Joint work with Adebayo, Gilmer, Goodfellow, Hardt, [NIPS 18]

TCAVs: Testing with concept activation vectors

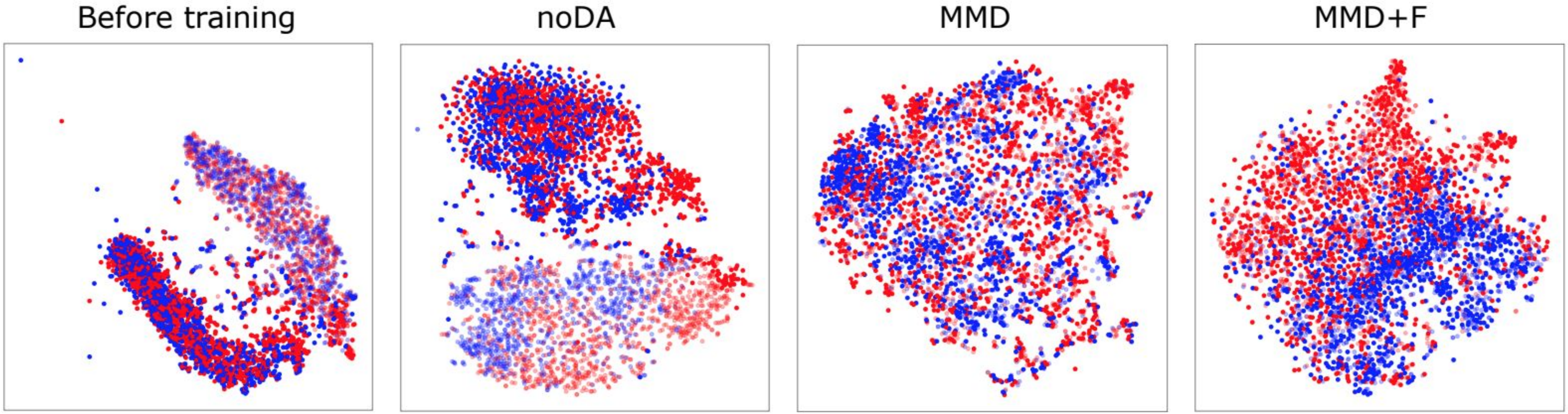
"[After the fact,] CAVs are learned by training a linear classifier to distinguish between the activations produced by a concept's examples and examples in any layer"

Interpretability beyond feature attribution: Kim+2018 https://arxiv.org/pdf/1711.11279.pdf, also https://www.youtube.com/watch?v=Ff-Dx79QEEY&ab_channel=MLconf

TCAVs: Testing with concept activation vectors

Interpretability beyond feature attribution: Kim+2018 https://arxiv.org/pdf/1711.11279.pdf,
also https://www.youtube.com/watch?v=Ff-Dx79QEEY&ab_channel=MLconf

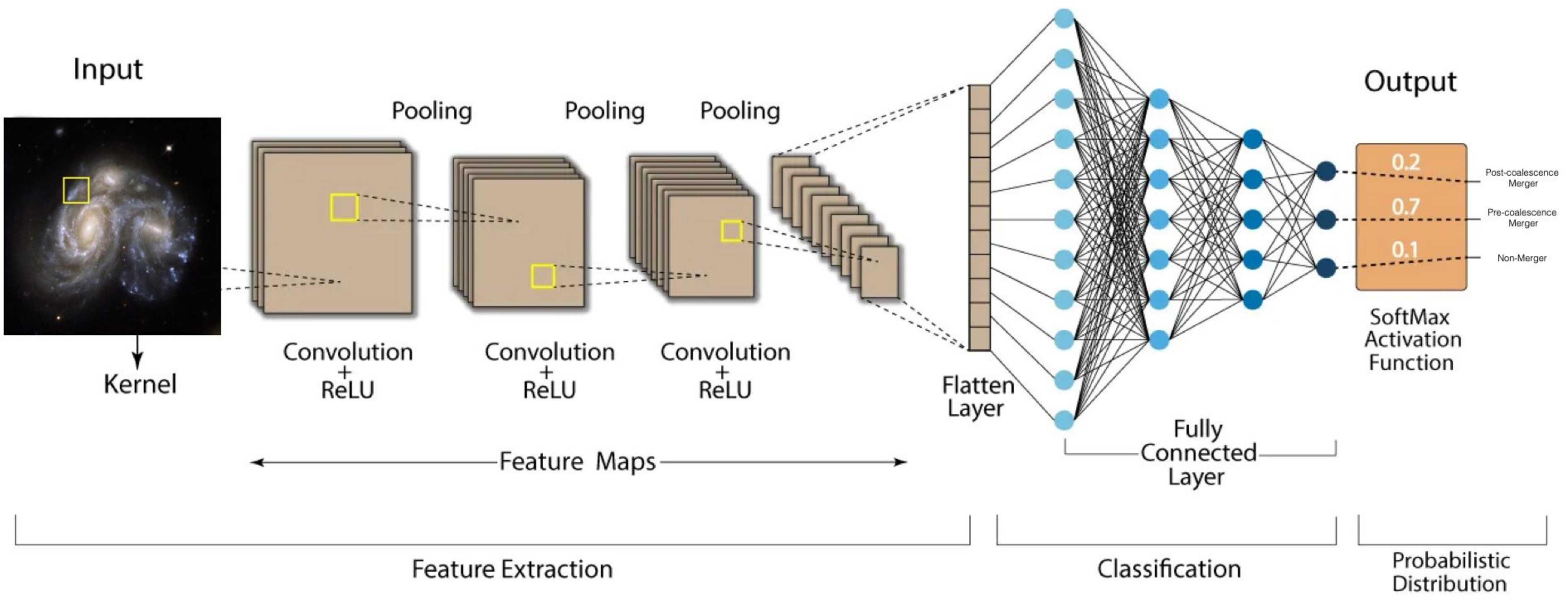
Domain adaptation finds invariant features between training and target data



red = mergers
transparent = Illustris,
solid = SDSS

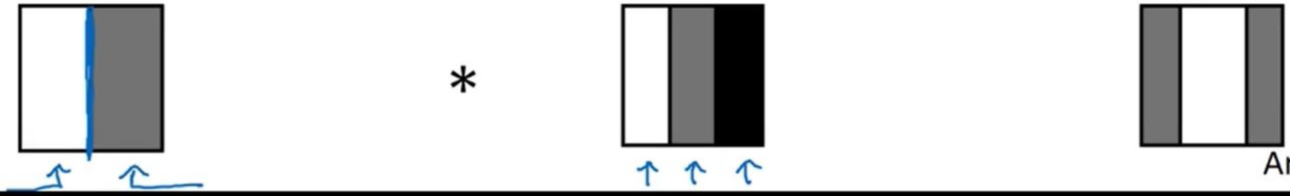
t-SNEs from Alexandra Ciprijnovic's 2021 paper --^

Convolution Neural Network (CNN)

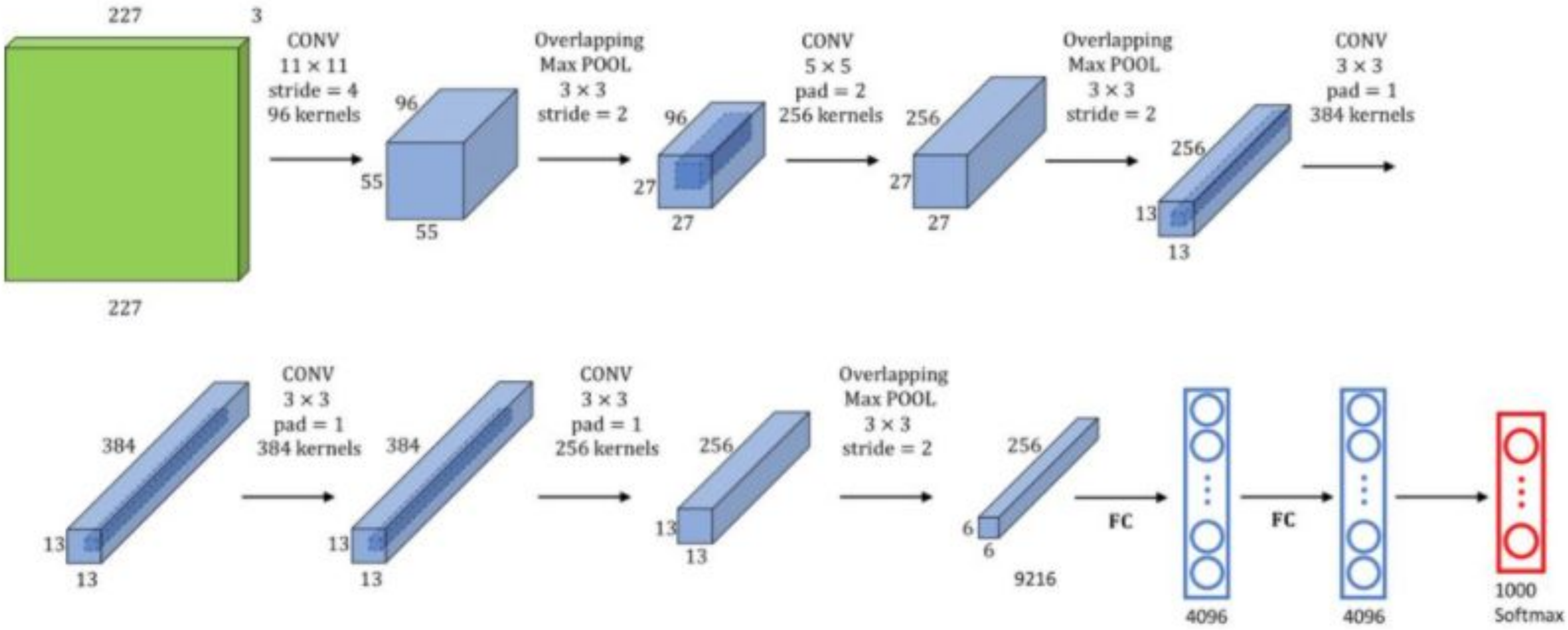


Vertical edge detection

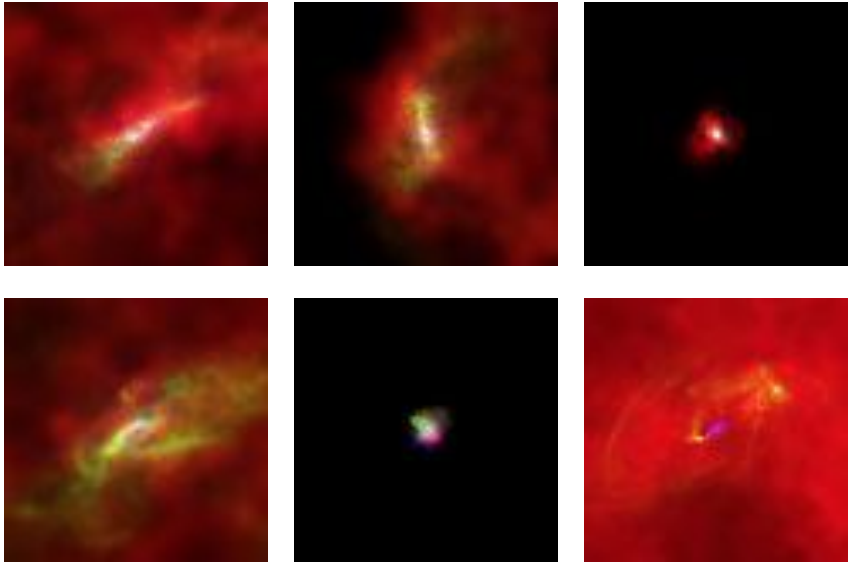
$$
\underbrace{\begin{bmatrix} 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \end{bmatrix}}_{6 \times 6}
*
\underbrace{\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}}_{3 \times 3}
=
\begin{bmatrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{bmatrix}
$$



Andrew Ng

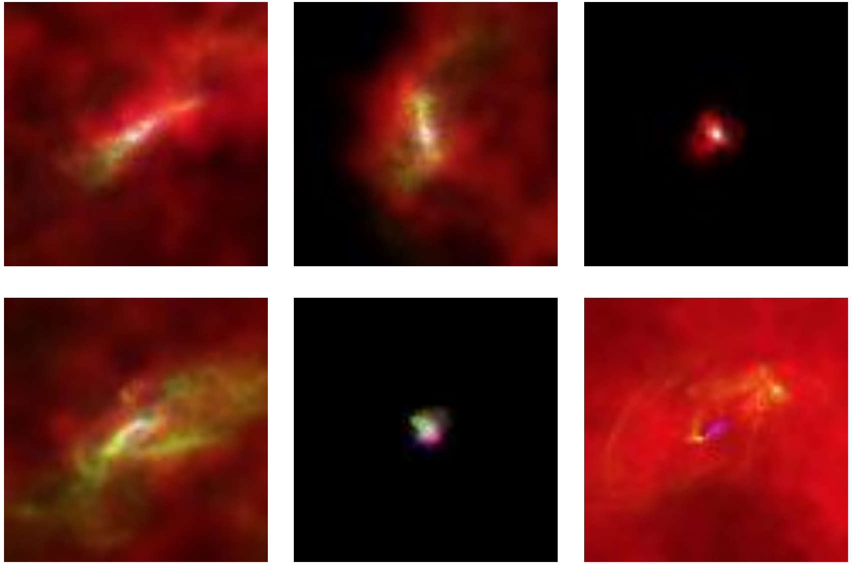

227
3
227
CONV
11 × 11
stride = 4
96 kernels
96
55
55
Overlapping
Max POOL
3 × 3
stride = 2
96
27
27
CONV
5 × 5
pad = 2
256 kernels
256
27
27
Overlapping
Max POOL
3 × 3
stride = 2
256
13
13
CONV
3 × 3
pad = 1
384 kernels
384
13
13
CONV
3 × 3
pad = 1
384 kernels
384
13
13
CONV
3 × 3
pad = 1
256 kernels
256
13
13
Overlapping
Max POOL
3 × 3
stride = 2
256
6
6
9216
FC
4096
FC
4096
1000
Softmax

Transfer learning is an exciting option



Options: TNG100 (8 times the volume)

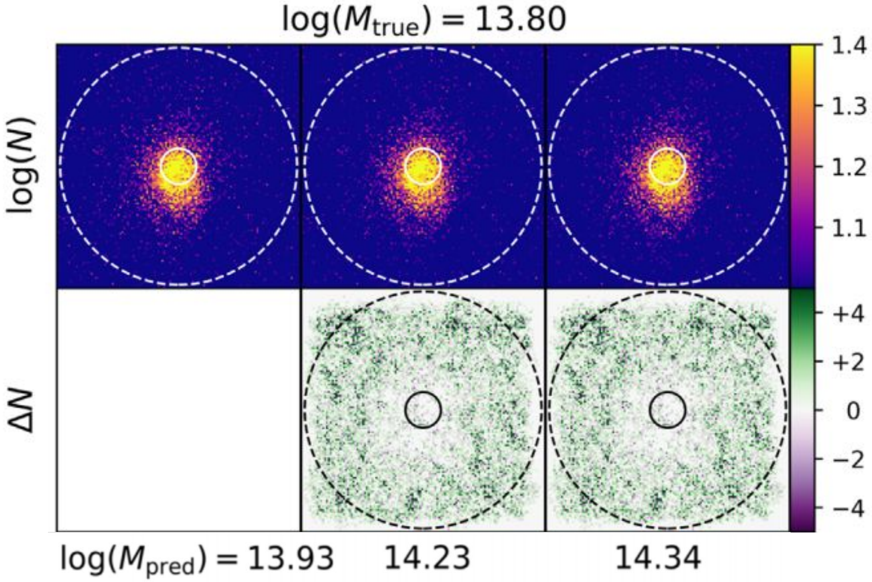
Transfer learning is an exciting option



Options: TNG100 (8 times the volume) or dogs and cats!!

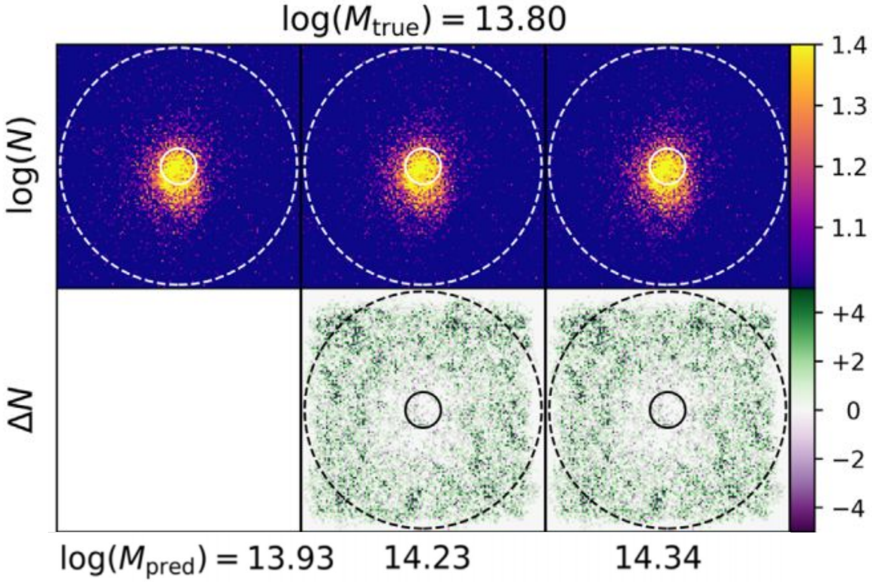
How do we untangle the CNN's decisions?

Saliency methods - e.g., Ntampaka+2018 use Google DeepDream to compute the gradient of the output



How do we untangle the CNN's decisions?

Saliency methods - e.g., Ntampaka+2018 use Google DeepDream to compute the gradient of the output



However, saliency maps can be misleading (Adebayo+2018)

Apparently there’s a hello kitty cafe